SPECIAL ISSUE ON

BIOTECHNOLOGY & HUMAN SYSTEMS

www.ll.mit.edu

On the Cover

Knowledge of the structure of neural networks is critical for improving the understanding of how the human brain works. Lincoln Laboratory and MIT researchers are developing a large-scale automated neuron-tracing capability to enable fast, accurate reconstruction of neural connectivity from high-throughput microscopic brain imagery. Using axon data collected by the CLARITY tissue-clearing technique, which allows acquisition of volumetric images of brain tissue at cellular resolutions, the team demonstrated an artificial intelligence–based high-performance pipeline that rapidly traces millions of axons. The cover image shows automatically traced axons in the densely packed hippocampus region of a mouse brain. The team is working on refining the automated neuron-tracing pipeline that ultimately will allow neuroscientists to gain insight into the underlying neural mechanism of brain disorders, thus facilitating therapeutic discovery.

Features

Letter from
the Director

In 1995, Lincoln Laboratory introduced into its portfolio several programs that applied our expertise in sensors and signal processing to the detection and identification of chemical and biological agents in the environment. Since then, our biology-related efforts have expanded into projects involving systems that not only help protect the environment but also enhance human health and performance. This issue of the *Lincoln Laboratory Journal* highlights recent and ongoing biotechnology work, representing research and development ranging from the creation of systems that sense individuals' physiological and neurological status to the investigation of innovative synthetic biology. These articles represent some of our more mature biotechnology projects; we continue to explore new applications.

As this special issue of the *Lincoln Laboratory Journal* goes to press, many parts of the United States and the world are emerging from stay-at-home guidance that has been one of the first steps to bringing the coronavirus pandemic under control. Reflecting on the human and economic impacts of the pandemic to the nation, we see research and development in biology, biotechnology, and human systems as a critically important part of national and global security. This issue provides a view of our commitment to an area of research and development that is growing in importance to improving the human condition on many fronts.

For the past decade, the infectious disease and public health community has warned the world of an increased risk of pandemics as a result of the growing human population and its encroachment on natural habitats, the interconnectedness of global travel and economies, and climate change expanding new disease vectors. The world is now experiencing the worst pandemic in a hundred years for all these and many other reasons. As we are in the midst of this crisis, we see that advanced technology is playing a key role in the response, from new molecular diagnostics and sophisticated computer modeling to rapid design and development of vaccines. It is clear that many gaps remain in pandemic preparedness and response. New technologies could enable an integrated system that will improve the ability of our nation and the world to effectively respond to potentially more dire future pandemics.

An effective, integrated pandemic response system would focus on four main objectives: surveillance, containment, mitigation, and eradication of the disease. The system's overall goal is to minimize the number of infected people while causing the least amount of disruption to society. This goal is easiest to reach when we can stay ahead of an outbreak, and a system that meets this goal requires several components:

- An improved global early-warning system that finds and characterizes new pathogens,
- Widespread, simple, and affordable diagnostic tests that enable the public health community to find and contain outbreaks,
- Efficient and effective contact tracing,
- Rapid drug repurposing to treat a new disease,
- Rapid new therapeutic and vaccine design, development, testing, manufacturing, and distribution,
- Improved health care system situational awareness that enables efficient resource allocation and patient care,
- Improved epidemiological and economic modeling to predict the costs and benefits of intervention policy options,
- Methods of understanding and predicting human behavior under various policy scenarios.

To advance the research and development in many of these areas, the Laboratory is planning to establish a new Biotechnology and Human Systems Division in the next couple of months. The Laboratory's long-standing expertise in systems analysis, sensing, and signal processing, and its more recent experience in biotechnology and artificial intelligence (AI) can provide significant contributions to all of the needed components. We have already begun new programs in some of these areas, leveraging AI for rapid drug discovery, automating contact tracing, and improving analysis of the medical system needs. As we look forward to transitioning new technologies to commercial companies and national health protection centers, we hope to support the nation with capabilities to better anticipate and respond to the next pandemic.

Eric D. Evans
Director, MIT Lincoln Laboratory

History of Lincoln Laboratory's Bioengineering Research

Edward C. Wack and Jeffrey S. Palmer

Lincoln Laboratory researchers began applying their considerable expertise in systems analysis, sensors, and signal processing to chemical and biological defense applications as early as 1995. This expertise, plus newer capabilities in synthetic biology and brain and cognitive sciences, has since been directed to include research into biotechnological solutions for identifying and remediating physiological, neurological, and cognitive problems impacting the health and performance of U.S. warfighters, veterans, and civilians.

>> **Since 1951, MIT Lincoln Laboratory has** remained at the forefront of technology development in support of national security. From the first integrated air defense system, the Semi-Automatic Ground Environment [1], that required advances in radars, communications, computing, and information theory to recent pioneering work in satellite optical communications [2], Lincoln Laboratory has developed cutting-edge scientific and technological capabilities and applied them to solve pressing national security problems.

As power generation and mechanization in the 19th century, and microelectronics technology and communications in the 20th century, created a host of new opportunities and challenges, the life sciences are poised to do the same in the 21st century [3]. The ability to read and write the fundamental code of life, the DNA letters of the genome, continues to increase exponentially. Understanding of the brain—its structure and function, injuries, and disease states—and the way this biological organ gives rise to the mind is accelerating each year. As scientists better understand the inner workings of life at all scales, they are able to cure diseases and repair injuries, and maintain optimal health and performance via the application of new technology. The human body attempts to keep itself in balance when confronted by changes in the environment or by injuries. From regulating internal temperature to blood glucose levels, the multiple physiological control systems strive to maintain homeostatic balance. When the bounds of the control system are exceeded, humans experience dysfunction, injury, or disease.

Technology enables an outer control loop to sense and respond to these "out of bounds" conditions. For example, continuous glucose monitors and insulin pumps are a technology-enabled solution to diabetes that keeps blood glucose levels in the proper physiological bounds. This integration of technologies to support human systems is enabled by the convergence of life sciences, physical sciences, microelectronics and nanoelectronics, and computer science. You will see examples of this technological convergence throughout this issue of the *Lincoln Laboratory Journal.* Our researchers at Lincoln Laboratory use speech signal processing with cutting-edge brain models of speech production to diagnose neurological conditions. We have developed novel wearable sensor technologies and detection algorithms to detect infection, prevent heat strain, and reduce lower-limb musculoskeletal injuries. And, we can create new genetic circuits to program cells to detect and destroy cancer cells in the body.

With the miniaturization and proliferation of technology, including biotechnology, we are working to achieve three important goals. We can intervene to restore health at multiple scales within the body, from the molecular and genomic level, to tissues and organs, to the entire human system, including social groups. Second, we can provide solutions that improve health not only in clinical or laboratory settings, but in regular living conditions at home and work. And finally, solutions can be tailored to personal needs, not based on population averages but on longitudinal data collected on individuals. The ability to measure and digitize human states, from the genome to blood chemistry and from physiology to cognitive states, enables the application of advanced computer science algorithms, artificial intelligence, and machine learning approaches to learn, analyze, and predict health states. It is an exciting time for developing bioengineering technology, and the convergence of these technical disciplines is accelerating innovation.

National Needs

It is fortunate that we are able to develop new approaches to solving health, performance, and security challenges as these challenges continue to grow in the complex environment in which we live. The U.S. military, in particular, faces multiple health challenges that affect soldiers' readiness and resilience, thereby impacting operational effectiveness. Such challenges create a fundamental national security risk if the United States is not able to field a physically, cognitively, and psychologically ready and resilient force. The threats come in many forms, internally and externally, as well as to individuals and populations.

Among the top concerns affecting the well-being of both soldiers and civilians are weapons of mass destruction, including novel and emerging biological threats. The Departments of Defense and Health and Human Services have done an outstanding job of developing sensors, medical countermeasures, and protective gear for traditional threats, such as *Bacillus anthracis* (the bacteria that causes anthrax disease), *Variola major* (smallpox), and *Yersinia pestis* (plague). However, the same biotechnology that allows genome editing to cure disease can be used to create new pathogen threats for which the world has no sensors or medical countermeasures. Novel sensors and rapid medical countermeasures are urgently needed to combat these novel threats [4]. Two articles in this *Journal* ("The Impact of Host-Based Early Warning on Disease Outbreaks" and "Synthetic Biology") describe work that begins to address these challenges. In fact, as this issue of the *Lincoln Laboratory Journal* goes to press, we are in the midst of refocusing multiple research efforts to support the nation's COVID-19 response. These initiatives include host-based early warning of infection, the use of speech to track disease symptoms, and the application of synthetic biology and bioinformatic tools to discover potential medical treatments.

Members of the U.S. services operate in challenging environments of high altitudes, extreme heat and cold, and increasingly in dense, urban megacities. These environments not only challenge service members' physical and cognitive capabilities, they also potentially present opportunities for servicepersons to confront near-peer adversaries who may contest the freedom of U.S. services to maneuver in the air, land, and sea and their ability to communicate. Such complex and dangerous environments have dramatic implications for battlefield medical care for those injured in combat. There is an expectation that the U.S. military will need to care for larger numbers of casualties and for longer time periods than were seen in the past 40 years [5–7]. No longer will U.S. field doctors be able to medivac the injured to medical treatment facilities within the "golden hour," the short period of time to

sustain life until trauma surgeons can repair injuries. The military will require prolonged field care to sustain life and treat injuries, possibly longer than 24 hours, before medical evacuation is possible. This issue of the *Lincoln Laboratory Journal* presents research on technologies that will assist battlefield medicine ("Noncontact Laser Ultrasound for Biomedical Imaging Applications").

Even before deployment and potential battlefield injuries, U.S. service members perform rigorous training to master their operational roles. They enter service with a range of strengths and weaknesses that training raises and levels. However, some risk factors, such as genetic determinants of bone health or preexisting psychological trauma, can be exacerbated in training, reducing an individual's overall readiness. Quantifying baseline performance and tailoring nutrition, training, and support to maximize individual performance without injury will be necessary so that the United States can continue to have the most capable military in the world. Lincoln Laboratory is developing technologies to prevent injuries in training environments (see "Open Body Area Network Physiological Status Monitor," "Model and Personal Sensor for Metabolic Tracking and Optimization," "Biomechanical Sensing and Algorithms," and "Understanding Noise-Induced Auditory Damage").

The military's future operational environment will include humans teaming with autonomous systems. This new paradigm will potentially revolutionize the character of warfare, and the implications for individual performance are poorly understood. What are the optimal forms of communication between humans and machines? How is understanding and trust built in these teams? How do the series of interdependent tasks for complex missions get assigned and dynamically reassigned across team members? These difficult questions require that we gain a better understanding of how to measure and analyze human and machine physical and cognitive states, and develop models by which to characterize and predict individual and team performance (see "Fundamental Brain Research").

Service members work in complex human teams as well as with autonomous systems. Being able to measure and improve team dynamics will also be critical for future operational success. Social systems extend beyond operational settings into garrison, home, and civilian life. Today, U.S. adversaries are attempting to undermine civil and social society via disinformation operations. A better understanding of how our minds can be trained to detect and reject such attempts is important for national unity and citizenship.

Finally, many of these same threats affect veteran and civilian populations: infectious disease, trauma injuries, psychological health, and disinformation on social media platforms. The technology solutions developed for the military will have broad applicability in the homeland for civil society.

The Lincoln Laboratory Approach

Over the decades, Lincoln Laboratory has made multiple, important contributions to biomedical technology development. We were pioneering codevelopers (with MIT professors) of optical coherence tomography, known as OCT, a critical diagnostic tool used in ophthalmology, cardiology, and upper gastrointestinal endoscopy [8]. The Laboratory also developed RF-based thermotherapy for cancer therapy [9]. Early work in DNA microarrays utilized the Laboratory's microelectronics fabrication capability to build some of the earliest gene chips. Each of these technologies was successfully transitioned to commercial application via startup companies that spun out of Lincoln Laboratory.

In 2009, Lincoln Laboratory took a step back to understand what role it might play in the bioengineering and biomedical space in support of its national security mission. A panel of Boston-area luminaries in the medical and life science arenas was convened, and the study team mapped out a number of areas in which Lincoln Laboratory had unique capability combined with a strong mission pull. In 2010, the Laboratory began investing in internally funded research in the applied biomedical sciences and in 2012 formed a new group, the Bioengineering Systems and Technologies Group (recently renamed Human Health and Performance Systems), that joined the Chemical and Biological Defense Systems Group in the emerging human systems and biotechnologies mission area. In August 2019, the Laboratory created a new Biological and Chemical Technologies Group to expand research in these areas.

The articles in this issue of the *Lincoln Laboratory Journal* highlight some of the technical areas in which the Laboratory is working and the end-user applications of that technology. Building on decades of experience,

we have been developing advanced sensors, algorithms, and systems, at all scales, to measure, model, and modify humans. These technologies are used to diagnose neurological conditions, quantify cognitive load, detect early onset of infections, measure biomechanical loads, and develop cell-based sensors to detect hazards in the environment and in the body. We work with civilian end users, military personnel, and clinicians in their respective environments to understand their critical and unique challenges. We conduct systems analysis and develop modeling approaches to identify the key variables through which advanced technology could enable orders of magnitude improvements in their missions. And, we transition the technologies to industry to enable the government to buy them, to improve the domestic economy, and ultimately to improve human well-being across our nation.

In the decade since we began to focus on this area, Lincoln Laboratory has won seven R&D 100 Awards for innovation in biomedical systems, has filed or been awarded 15 patents for its biotechnologies, and has seen one new biotech company spun off. Going forward, we will continue to solve current critical challenges in health, performance, and security in partnership with the government, academia, industry, and hospitals. We will continue to build unique infrastructure and technology to enable cutting-edge research and development that is both accessible and adaptable for tomorrow's needs.

Welcome to this special edition of the *Lincoln Laboratory Journal*. We hope you enjoy the articles and gain new insights into solving some of our nation's biggest challenges in health and performance. As our nation combats the current COVID-19 pandemic and its effects on our security, Lincoln Laboratory, with many other research organizations, is rapidly responding to urgent military and civilian needs. It is through our united effort we will innovate the solutions to bring the crisis to an end. ∎

References

1. "SAGE: Semi-Automatic Ground Environment," MIT Lincoln Laboratory website, https://www.ll.mit.edu/about/history/sage-semi-automatic-ground-environment-air-defense-system, accessed 24 July 2019.

2. "Lunar Laser Communication Demonstration: NASA's First Space Laser Communication System Demonstration," NASA facts, https://www.nasa.gov/sites/default/files/llcdfactsheet.final_.web_.pdf, accessed 24 July 2019.

3. P. Sharp, T. Jacks, and S. Hockfield, eds., "Convergence: The Future of Health," 2016 Report to Convergence Revolution. Cambridge: MIT, June 2016, available at http://www.convergencerevolution.net/s/Convergence-The-Future-of-Health-2016-Report-55pf.pdf, accessed 24 July 2019.

4. National Biodefense Strategy. Washington, D.C.: White House, 2018, available at https://www.whitehouse.gov/wp-content/uploads/2018/09/National-Biodefense-Strategy.pdf, accessed 24 July 2019.

5. M. Kinney, "Evolving 'Golden Hour' Care to Mitigate the Treatment Gap," *Combat & Casualty Care*, spring 2018, pp. 4–6.

6. "Multi-Domain Battle: Combined Arms for the 21st Century," U.S. Army–Marine Corps White Paper, 18 January 2017, available at https://ccc.amedd.army.mil/PolicyPositions/Multi-Domain%20Battle%20-%20Combined%20Arms%20for%20the%2021st%20Century.pdf, accessed 24 July 2019.

7. A.S. Linde, "The Need for Pre-hospital Simulation Training Platforms in Battlefield Medicine," *Journal of Trauma and Rehabilitation*, vol. 1, no. 1, 2018.

8. J. Fujimoto and E. Swanson, "The Development, Commercialization, and Impact of Optical Coherence Tomography," *Investigative Ophthalmology and Visual Science*, vol. 57, no. 9, 2016.

9. A.J. Fenn, *Adaptive Phase Array Thermotherapy for Cancer*. Norwood, Mass.: Artech House, 2009.

About the Authors

Edward C. Wack is the assistant head of the Homeland Protection and Air Traffic Control Division at Lincoln Laboratory. In this role, he shares responsibility for research, development, evaluation, and technology transfer of advanced technologies and systems for chemical and biological defense, bioengineering, and biomedical systems. Prior to holding this position, he was the leader of the Bioengineering Systems and Technologies Group, which focused on innovative advanced technology programs in biodefense, forensics, and biomedical research. He was also Director of Future Acquisition at the Joint Program Executive Office for Chemical and Biological Defense (JPEO-CBD) within the Department of Defense. At JPEO-CBD, he was responsible for leading the office's future technology strategy and coordinating that strategy with the other members of the chemical and biological defense program, civilian government

organizations, and international partners. Before joining the JPEO-CBD, he spent 13 years at Lincoln Laboratory, where, as an assistant leader of the Sensor Systems and Applications Group, he led a team working on standoff sensing, advanced detection algorithms, and system architectures. He has also been involved in various aspects of satellite remote sensing programs, including system architectures, sensor designs, sensor calibration, and requirements analysis and definition. He holds a bachelor's degree in mathematics from the College of the Holy Cross and a master's degree in bioinformatics from Brandeis University.

Jeffrey S. Palmer is the leader of the Human Health and Performance Systems Group at Lincoln Laboratory. He has oversight of multiple research programs that focus on health, human performance, objective neurocognitive analytics, and biosensing via wearable, ingestible, and implantable devices. In 2010, he helped to create the first Army War College Fellowship at Lincoln Laboratory and the MIT Security Studies Program. He has given presentations at international conferences and authored book chapters and technical articles on DNA biometrics and forensics, biomechanics, cell biology, materials science, soldier nanotechnology, biological-chemical defense, polymer science, high-energy lasers, microelectronics packaging, wearable biomedical sensing in extreme environments, and neurocognitive technologies. He has served on editorial boards for journals in biomechanics, molecular science, biomedical informatics, and biosensors. He has chaired technical conferences for the National Science Foundation, Department of Homeland Security, and the IEEE. Currently, he is the vice chair (and chair-elect) of the IEEE Engineering in Medicine and Biology Society's (EMBS) Technical Committee on Wearable Biomedical Sensors and Systems and the EMBS conference editorial board for tissue engineering and biomaterials. In addition, he has served as an advisor on senior military studies of enhancing health and performance, and led a multi-agency U.S. government effort to develop automated rapid human DNA analysis capabilities for field biometrics and forensics applications. Prior to working at Lincoln Laboratory, he worked at research laboratories at IBM and GE, and at the Physical Sciences Laboratory at New Mexico State University. He holds a bachelor's degree with a minor in mathematics from New Mexico State University, a master's degree from Rensselaer Polytechnic Institute, and a doctorate with a minor in bioengineering from MIT, all with majors in mechanical engineering.

Lab *Notes*

BIOMEDICAL DEVICES

An Instrument to Transform Back Surgery

The design of a new laser-based scope could offer a less invasive, more precise way to remove tissue that causes back pain and immobility

---

A casual conversation between friends sparked the development of an innovative tool that could change the way doctors perform surgery to alleviate lumbar spinal stenosis, a condition resulting from the narrowing of the spinal canal. At a backyard cookout, Matthew Johnson, a technical staff member at Lincoln Laboratory, was talking with childhood friend Patrick Codd, then a neurosurgeon at Massachusetts General Hospital in Boston. Codd was explaining the difficulty of accessing the spinal column to remove overgrown connective tissue that compresses the spinal cord and nerve roots, causing severe back and leg pain and weakness, even paralysis.

"There's got to be a better way," Codd commented.

Johnson, an engineer who has always been intrigued by solving problems, began to think about that better way. "I had zero background in anatomy, medical devices, or lasers," Johnson said. "I was working on electrical engineering solutions for a variety of systems, including lasers." But the problem gnawed at him. Could a laser beam excise, or ablate in medical terminology, the obstructive connective tissue? Doctors already were using lasers to remove damaged soft tissue or kidney stones. What type of instrument could navigate the spinal canal and direct a laser to safely and efficiently remove the tissue?

Johnson sought advice from colleagues with experience in lasers and anatomy, and he worked closely with Codd and former Laboratory staffer Wes Hill to design a laser-based tool that could access the spinal column and then ablate the compressive tissue. He took a proposal for this tool to the Laboratory's Advanced Concepts Committee, which grants internal funding for experimental research. The committee approved the project that could lead to improved treatment for the 8.3 million Americans who suffer from lumbar spinal stenosis, many of whom are military personnel whose training and active-duty assignments have often triggered the onset of the condition as early as age 40.

Tara Boettcher of the Laboratory's Human Health and Performance Systems Group was Johnson's go-to person for the anatomical aspects of the tool. "I handled all biology aspects of the program to execute it here at MIT Lincoln Lab," Boettcher said. "Matt and I adapted a room so that lasers and biology could co-exist following all safety regulations to perform the experiments."

The standard initial treatment for lumbar spinal stenosis is physical therapy. If physical therapy is not effective, the next step is often an epidural steroid injection administered through a needle into the space around the spinal cord. This injection of an anti-inflammatory medicine is usually prescribed after the stenosis has progressed enough to cause the patient difficulty walking or some neurologic damage. Injections are generally administered every few months until they are no longer effective. At this point, surgery is often the next step.

"Surgical decompression is reasonably a measure of last resort," Johnson explained. "Though only a small amount of compressive tissue needs to be removed, surgery involves significant bone and tissue destruction, general anesthesia, and a long recovery time, and sometimes it even necessitates spinal fusion."

To develop a surgical tool capable of ablating compressive

tissue in the spinal canal without destroying healthy tissue, the researchers at Lincoln Laboratory collaborated with Codd on the practicalities of the laser sacroscope. Their goal was to shift the treatment model toward early surgical intervention that would relieve pain, preempt further physiological damage, require less extensive surgery and rehabilitation, and ultimately reduce hospital costs.

The prototype laser sacrascope was built around a commercially available ureteroscope. The sacrascope's features are designed for the three primary phases in surgery to remove compressive tissue: access the spinal canal, identify the compressive tissue, and remove it.

With an outer diameter of roughly 3 millimeters, the sacrascope can access the spinal canal through a 2-centimeter incision in the skin that covers a naturally occurring hole in the sacrum (a bone at the base of the spine) called the sacral hiatus. Once the device has entered the sacral hiatus, the surgeon can gently guide the device up the spine along the dorsal (upper) wall of the spinal canal. The sacrascope's shape conforms to the curvature of the spinal canal, allowing the device to smoothly track along the dorsal canal wall and not get caught on any nerve roots.

"To allow the surgeon to positively identify the compressive tissue, we wanted the sacrascope to provide high-fidelity video imagery," Johnson said. The sacrascope uses a sensor capable of advanced image processing. The sensor detects the relative

The sacrascope's features are designed for the three primary phases in surgery to remove compressive tissue: access the spinal canal, identify the compressive tissue, and remove it.

brightness and darkness of tissue that is close to (bright) or farther from (dark) the device's tip to create an accurate video image.

"Because the sacrascope is introduced into a compressed anatomy, we developed two petals that, when actuated, hinge laterally away from the side of the distal tip of the device while the device face concurrently slides backward. This mechanical action reveals a clear area for direct visualization and a safe surgical area, with the spinal cord and nerve roots protected from the laser fiber by the lower petal," Johnson explained.

To remove the compressive tissue, two concentric nitinol (a nickel titanium alloy) tubes advance from the working channel of the sacrascope. The straight, stiff outer tube carries a laser fiber and has a predefined radius of curvature. By sliding the inner tube relative to the outer tube and rotating the tube assembly, the laser fiber can be steered anywhere within the field of view of the imaging sensor.

For control of the inner and outer tubes, the Lincoln Laboratory engineers developed a unique

mechanical handle at the far end of the sacrascope. The handle is designed so that relatively large movements by the operator result in very fine, controlled movement of the laser fiber within the spinal canal. This handle enables precise targeting of only the intended tissue, without the need to reposition the entire device for each new target area.

Although lasers have been commonly used for some procedures on humans, their effect on tough, rubbery tissue like that compressed in stenosis, the ligamentum flavum, was not well documented. Therefore, the sacrascope development team devoted substantial effort to performing experiments to understand the properties of ligamentum flavum and the impact of laser beams on that tissue type.

One consideration in using lasers is the effect of the beam's heat on tissue. "We wanted to minimize damage to tissue adjacent to the tissue being ablated," Johnson said. "We looked at the absorption characteristics of the ligamentum flavum and selected two promising lasers; one commonly used in medicine, and one that isn't currently used in medicine but may offer a large advantage in terms of precision."

The team then tested the lasers on ligamentum flavum obtained from the spine of a pig. To evaluate how lasers affected "living" tissue, the experiments were done on tissue taken no more than four hours after the animal's death and were completed within 10 hours after the tissue was acquired. The

researchers irradiated the tissue to assess the amount of energy needed to ablate a specified area of tissue and the amount of thermal damage to adjacent tissue.

The Ho:YAG laser, which irradiated the ligamentum flavum through a 600-nanometer-diameter fiber-optic cable, rapidly heated, thus ablated, the tissue. This laser is already used for surgical procedures, and its beam is easily transmitted through a standard multimode fiber optic. However, the Ho:YAG laser caused significant thermal necrosis and charring of the adjacent tissue.

The frequency-quadrupled Nd:YAG laser, which focused the laser beam through an aperture and lens, achieved a clean ablation with minimal heating (thus charring) of the adjacent tissue. However, this laser wavelength is difficult to transmit at sufficient power levels through traditional fiber optics.

Prior research on lasers for surgery to mitigate spinal stenosis has not produced a usable tool. "Three things have made other approaches unsuccessful," Johnson explained. "The doctors could not get an endoscope into the spinal canal well enough to see what they were doing. They were unable to steer the laser successfully, and the tissue got charred."

The sacrascope uses a camera that is small and produces high-quality imagery, and its design allows for the laser to be precisely steered to the area of interest. The final challenge is figuring out the best type of laser to use. The Lincoln Laboratory researchers are no longer directly involved in
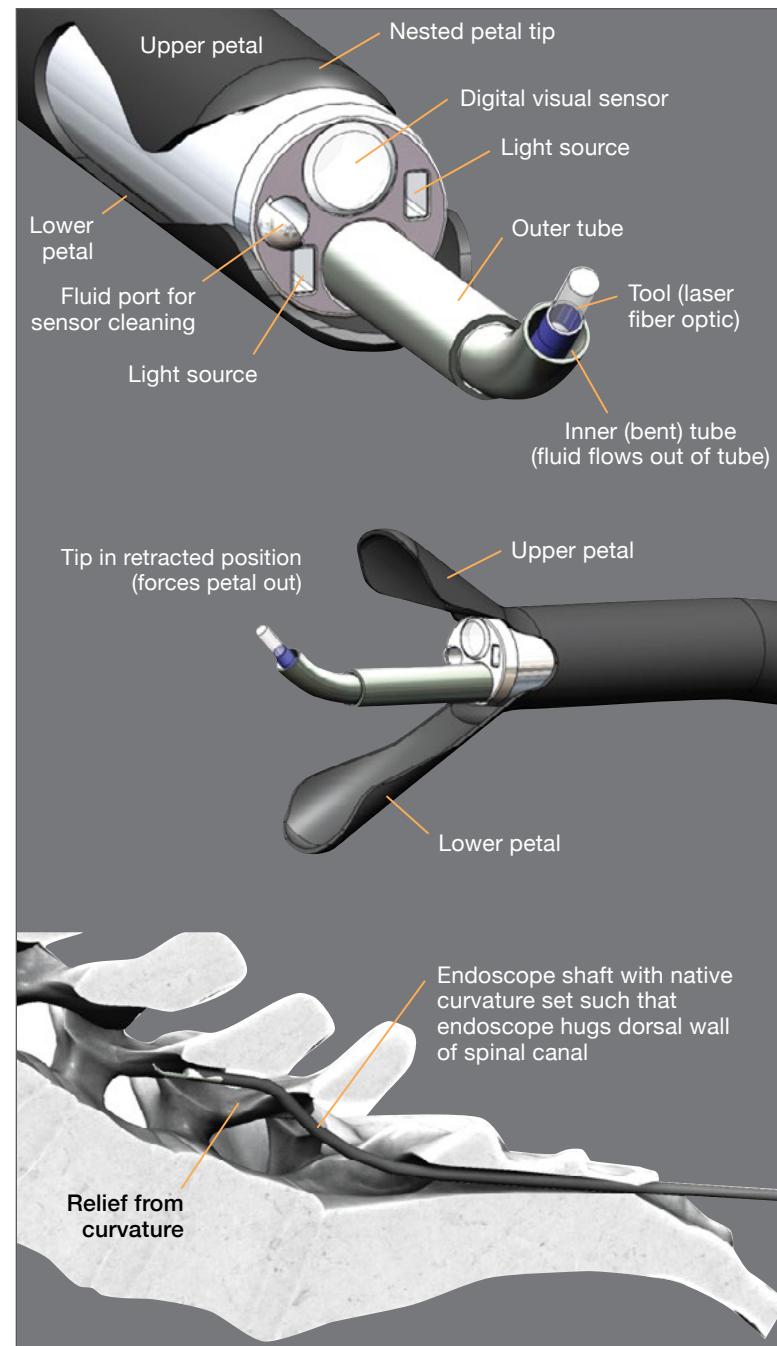
An endoscope was modified to build the laser sacrascope (top left); the callouts point out the modifications. In the middle is shown the action of the petals that deploy from the side of the sacrascope's face, clearing a safe surgical field. At bottom, the laser sacrascope is inserted into the spinal canal through the sacral hiatus and ascends the spinal canal along the dorsal wall.

developing the sacrascope, but they have filed patents on the technology and hope a biomedical device company will pick up the project

to investigate that best laser and develop a marketable sacrascope.

Boettcher sees the development of the tool as a valuable

investigation. "One of the cool things about MIT Lincoln Lab is the ability to combine experts from many different fields to work on ideas that can be prototyped and someday have an application," she said. "We published a paper on our work in *Lasers in Surgery and Medicine*, and the technology has been licensed. I hope to see it in the clinic someday soon."

Another outcome of the sacrascope project is the creation of a new bioengineering facility within Lincoln Laboratory. Boettcher explained that the space she and Johnson adapted for the sacrascope experiments "has now been made more official and permanent, called the BEAMM Lab."

The Biophotonic, Electric, Acoustic, and Magnetic Measurement (BEAMM) Lab is a facility that will allow staff to apply the Laboratory's capabilities in optics, lasers, radio frequency, imaging, and image processing to investigate how biological tissue reacts to the application of different energy types.

"The BEAMM Lab is a unique resource that will enable multimodal, multispectral characterization of biological materials," said Catherine Cabrera, leader of the Biological and Chemical Technologies Group. "Biomedical imaging has been identified as a key area in which the Laboratory could have a significant impact by leveraging its resources."

Fluorescence Imaging for Surgery

A near-infrared fluorescence imager helps surgeons quickly locate microtumors

---

Clinical treatment of advanced cancers involves removing a large amount of affected tissue. The surgeon must remove as much cancerous tissue as possible while at the same time preserving enough healthy tissue to provide a good clinical outcome, minimize patient trauma, and speed recovery. Affected tissue can present with tumors ranging in diameter from many centimeters (e.g., a golf ball) to submillimeter (e.g., a single poppy seed). Recent clinical studies indicate that removing submillimeter-sized "microtumors" can significantly increase post-operative survival. However, small tumors are extremely difficult for a surgeon to differentiate by eye from healthy tissue within a complex surgical field. This difficulty greatly impedes the detection and removal of the tumors.

The Second-Window Infrared Triband Fluorescence Imager (SWIFTI) enables surgeons to quickly scan the entire surgical field for tumors of any size and to accurately distinguish microtumors from healthy tissue.

SWIFTI takes advantage of unique fluorescent chemical "probes" that make the tumors light up on the SWIFTI video display. These chemical probes were developed by MIT researchers led by Angela Belcher, head of MIT's Department of Biological Engineering, and are made of single-walled carbon nanotubes (SWNTs) that bind to tumors but leave healthy tissue alone. When struck by a laser, the probes emit a fluorescent glow in the second near-infrared (NIR-II) wavelength band.

"To the best of our knowledge, no other intra-operative near-infrared fluorescence imagers currently operate in the NIR-II band with similar sensitivity, simple optical design, and precision for dual-mode fluorescence plus contextual imaging," said Andrew Siegel, who developed SWIFTI with fellow staff member Nandini Rajan. "SWIFTI was developed to explore the many advantages of NIR-II band fluorescence imaging for biomedical applications."

The first near-infrared (NIR-I) band, loosely defined as ranging from deep red at 650 nanometers out to 960 nanometers in the near-infrared, is the original biological "window" in tissue. However, this window actually extends beyond 1800 nanometers. In deference to the original biological window band, the longer wavelength region spanning from 950 nanometers to around 1800 nanometers is called the NIR-II band. In this band, there is deeper light penetration and less fluorescence generated by the tissue itself. However, to detect light at these wavelengths, special fluorophores and NIR-II-sensitive detectors are required.

Image across the entire NIR-II band

The SWIFTI fluorescence imager uses diffuse light from an 808-nanometer laser diode to excite fluorescence in single-walled carbon nanotubes within the tumor-selective probe molecule. Fluorescent emission, which can range from 1200 nanometers (nm) to 1700 nanometers, is then detected by an indium-gallium-arsenide (InGaAs) infrared (IR) camera. Spectral filters on the laser and visible light sources block any longwave emission that could be seen by the camera, to maximize tumor contrast. A spectral filter ahead of the IR camera blocks the relatively strong 808-nanometer light from being seen because the InGaAs detector material still exhibits some response at 808 nanometers.

Existing intra-operative fluorescence imaging systems operate in the NIR-I band. These systems all employ silicon-based charge-coupled-device or complementary metal–oxide semiconductor cameras, so they all rely on fluorophores with very narrow Stokes shifts (the shifts in wavelength between the excitation light and fluorescent emitted light). NIR-II window operation requires a different camera capable of detecting longer wavelengths. But it also allows the use of new fluorescent probes with much larger Stokes shifts, greatly simplifying the optical design. Also, since healthy tissues autofluoresce much less in the NIR-II band, the image contrast is very high, greatly improving tumor detection.

MIT's SWNT fluorophores combined with Lincoln Laboratory's SWIFTI imaging system exploit the many imaging advantages of the NIR-II band. The wide, approximately 800-nanometer Stokes shift provided by the SWNT probes allows the use of inexpensive optical filters with both good spectral blocking and broad acceptance angles. These filters reduce excitation light leakage and improve sensitivity because broader acceptance angles allow more light from the tissue to reach the camera.

As an added bonus, NIR-II wavelengths penetrate human tissue much deeper than visible light can, so tumors shadowed by a thin layer of tissue will still glow diffusely in SWIFTI video imagery, and many subsurface tumors can still be detected.

To use SWIFTI, the surgeon first applies the probe solution to all tissue in the surgical field and, after a short time, rinses away the excess unbound probe solution. When the surgeon depresses a footswitch, an infrared excitation laser at 808 nanometers illuminates all of the tissue in view of the camera. The tumors, now covered in fluorescent probes, re-emit some of this laser light within the NIR-II band. Healthy tissue, which has no fluorescent probes bound to it, appears dark on the display.

Lab Notes

Even small tumors glow brightly, distinctly recognizable against dark healthy tissue, making them much easier for the surgeon to both find and resect (remove) with confidence. Because the sharp, high-contrast imagery also reveals any unresected tumor, this image helps minimize the surgical margin (the healthy tissue surrounding the tumors that is also resected to ensure that the tumor has been completely removed), potentially improving patient outcome.

As the SWIFTI camera can still weakly detect light from the infrared laser used to excite the fluorescence, an optical filter is placed in front of the camera lens to block any wavelengths shorter than 830 nanometers. The result is a sharp, clear image in which all the surgeon sees is the bright fluorescent glow from the tumors against a dark, nearly clutter-free background.

In addition to the laser excitation source, SWIFTI also includes two more lights: a white LED light to illuminate the entire surgical field and a low-power 970-nanometer light source that only the NIR-II camera can see. Each source is separately dimmable and controlled by a footswitch.

The white LED light appears similar to a halogen surgical light, but it emits far less infrared light than incandescent lamps do, so it does not interfere with the fluorescent image. The in-band NIR-II light creates a grayscale image of the tissue that reveals surface glint (caused by moisture) and tissue contours similar to those seen in a visible monochrome image. Surgeons now have the freedom to operate from looking at the SWIFTI display screen. Using the footswitches, they can select a fluorescence view, a contextual grayscale view, or both simultaneously. They no longer need to switch their attention back and forth between the display screen and the surgical field, saving time and reducing fatigue.

"The innovative use of in-band illumination allows one camera to generate imagery that normally requires the input of two cameras in other systems, and it provides perfect spatial registration between grayscale and fluorescence imagery," Rajan said. This ability eliminates the need to maintain precise optical alignment or to remap the image in the digital domain, simplifying the design and reducing the cost of the system.

SWIFTI has been tested by surgeons and oncologists at Massachusetts General Hospital (MGH). Using the system on mice injected with the SWNT fluorescent probes, surgeons were able to remove tumors as small as 0.3 millimeters in diameter. Promisingly, the mice that were operated on using SWIFTI survived 40 percent longer than those who had tumors removed without the help of the system.
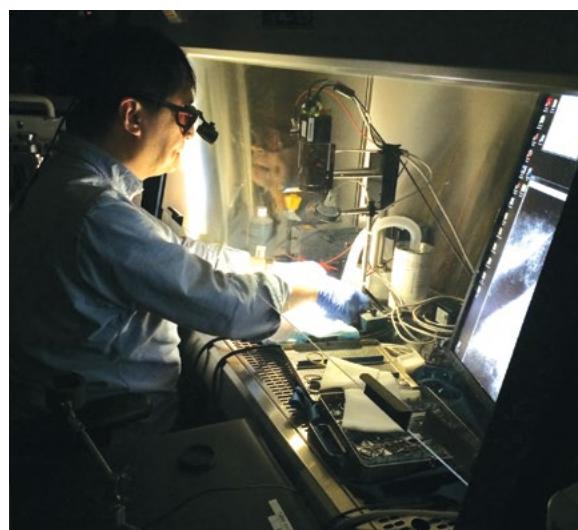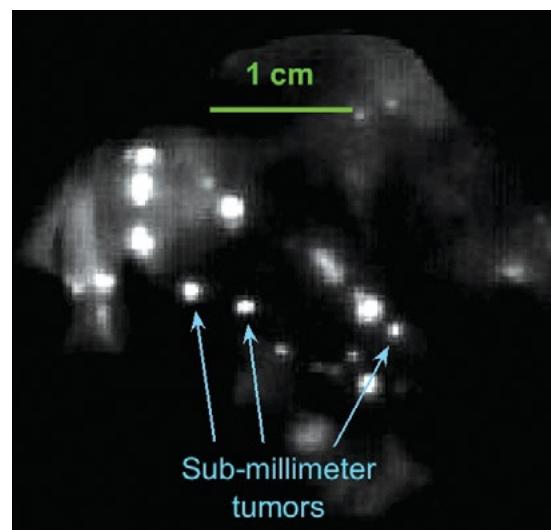
(a) (b)

The SWIFTI system facilitates the surgical resection of millimeter-scale tumors in mice (a). Even submillimeter tumors fluoresce brightly against the much darker healthy tissue, as shown in the fluorescence image of a mouse abdomen (b). The liver—visible as the lighter tissue region in the left side of this image—can be distinguished from the surrounding tissue by its slightly higher autofluorescence.

The MIT, MGH, and Lincoln Laboratory team has sought Federal Drug Administration approval for a Phase 1 clinical trial to test the imaging system for use on human patients. The system could be especially impactful for treating cancers that are difficult to diagnose in early stages, such as ovarian cancer. Seventy-five percent of the 250,000 new ovarian cancer cases diagnosed each year are in advanced stages, after a patient's abdomen is already riddled with many, and often tiny, tumors. SWIFTI could help surgeons resect many more of these tumors than can be removed today.

Diamond Sensors for Brain Imaging

A new detector uses a defect in diamond to sense faint magnetic fields emanating from the brain

When a neuron fires in your brain, the electrical current produces a magnetic field. Just as the Earth's magnetic field emanates from its core to outer space, the brain's magnetic fields ripple out past the skull, where they can be detected by sensors positioned around the head. Using these sensors, neuroscientists can estimate from which neurons the magnetic fields originated, allowing the scientists to pinpoint brain activity in real time.

This method of imaging the brain, called magnetoencephalography, or MEG, "is a phenomenal technology," said Danielle Braje, a physicist at Lincoln Laboratory, who added that MEG is used to diagnose post-traumatic stress disorder, localize the source of epilepsy, and study autism. Despite its clinical value, there are only a few dozen MEG brain scanners in the United States.

The sensors in today's MEG platforms require highly specialized facilities that cost millions of dollars to build and maintain. These platforms use superconducting quantum interference devices (SQUIDs) that operate at cryogenic temperatures and require constant

cooling. Moreover, clinical operation requires a magnetically shielded room, so that external magnetic fields, such as the Earth's, don't obscure the patient's brain signals.

Aiming to make MEG more accessible, a team of researchers from Lincoln Laboratory and MIT is developing a new sensor that could drastically cut operation costs while providing the same magnetic field sensitivity as SQUIDs.

The new sensor is made from diamond with a crystal structure optimized for magnetic field sensing. In nature, a "pure" diamond is composed of a lattice of carbon atoms with numerous defects and impurities. At the Laboratory, the team is growing
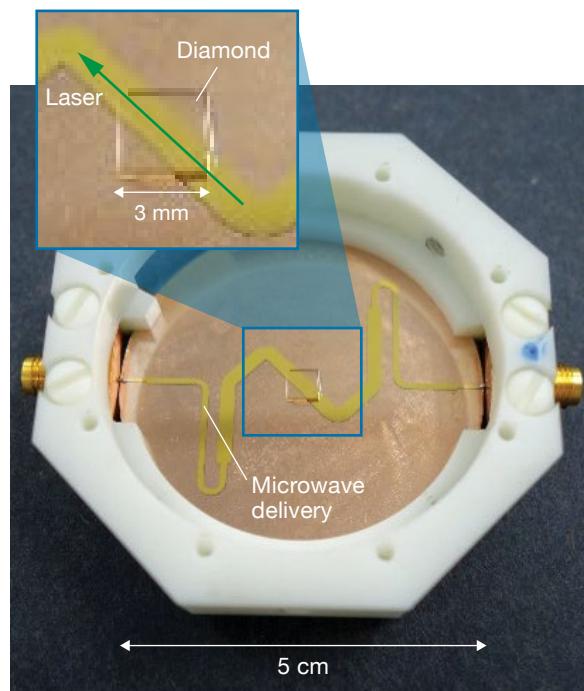
Green laser light is shone through the diamond, situated in the center of the sensor, to initialize the energy states of atoms in the diamond's nitrogen vacancy centers. These energy states transition to different levels in the presence of a magnetic field. By tuning microwave energy on resonance with these transitions, scientists can measure the transitions and therefore measure the magnetic field. This sensor would sit directly on a patient's head.

synthetic diamonds with purposeful defects. When a carbon atom is swapped for a nitrogen atom, an atom-like defect appears in the neighboring lattice site. "This is called a nitrogen vacancy defect or color center, and it is incredibly sensitive to magnetic fields," Braje said. A millimeter-sized diamond sensor contains millions of these nitrogen vacancy defect centers and can aggregate their signals for high-sensitivity detection.

Just as a compass will point north to align itself to a magnetic field, the energy levels of atoms in the nitrogen vacancy center will shift in response to an applied magnetic field. To measure this shift, the ensemble of atoms in a nitrogen vacancy are initialized into their lowest energy state, typically with a green laser. By tuning microwave energy on resonance with the nitrogen vacancy's energy transitions, an interferometric measurement can be made to deter-mine the shift of the energy levels, thus determining the magnetic field.

"One consistent challenge is engineering the sensor to have better sensitivity," said John Barry, who works in the Laboratory's Quantum Information and Integrated Nanosystems Group. The brain's magnetic fields are extremely tiny, about 100 million times weaker than the Earth's. The diamond sensor is expected to be sensitive enough to detect changes of about one billionth of Earth's field after only one second. "Of course, we are always trying to improve the quality of the diamonds that form the heart of the sensor," Barry said. Optimizing the
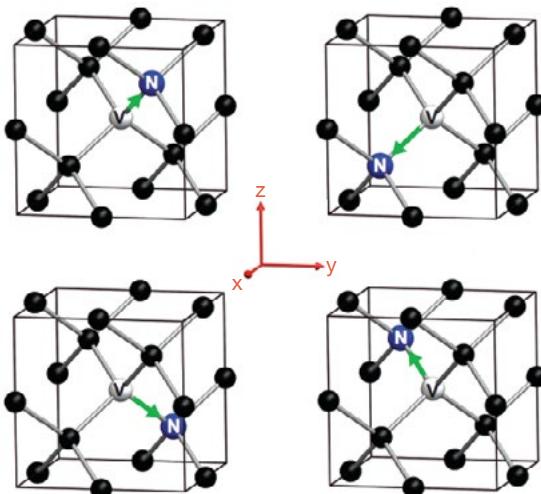
There are four possible orientations of the nitrogen-vacancy (NV) center defect in each unit cell, or the basic repeating structure, of the diamond crystal lattice. When an NV center in each orientation is used to sense a magnetic field, the direction of the magnetic field can be reconstructed.

type of carbon and the isotope of nitrogen that are used in the diamond is an important aspect of improving the diamond's quality. "Better diamonds make building a more sensitive sensor much easier," he added.

While its sensitivity to the strength of a magnetic field is expected to be comparable to that of a SQUID, the diamond sensor can actually describe a magnetic field in more detail because it provides a full vector measure-ment. A complete description of a magnetic field at any point in space requires two pieces of information—the magnetic field's strength and its direction relative to all three axes. A single SQUID measures field strength and only one directional axis of a magnetic field. A single diamond sensor can measure all three directional axes of a magnetic field because the sensor's nitrogen vacan-cies are arranged among four different orientations within the diamond's carbon lattice. When nitrogen vacancies in these

different orientations are used to sense a magnetic field, its direction can be reconstructed.

The diamond sensor also works at room temperature, in contrast to SQUIDs, which are only sensitive to magnetic fields at extremely cold, or cryogenic, temperatures. Because of this cryogenic requirement, the array of SQUIDs that surround a patient's head are built into a hard helmet under which the patient sits. "It's a one-size-fits-most solution," Braje said. "You have to hope your head isn't too big to not fit but not so small that the sensors are farther away from the skull. Only two MEG machines are tailored for children."

By contrast, the diamond sensor would rest directly on the person's head, strung in a net of hundreds that could conform to that person's head size and shape. This direct contact not only provides a better measurement, but it also allows measurements to be made without the need for a magnetic field–shielded room. "Since nitrogen vacancy sensors are not cryogenic like SQUIDs, they

can conceivably be spread out in a large 3D array covering the patient's head that can allow the background magnetic field to be measured and subtracted, negating the use for the shield," said Barry. "Although such an approach will work in principle, it is expected to be quite difficult to implement in practice."

Today, the team is collaborating with Massachusetts General Hospital (MGH) in their initial steps to put the sensor into practice. As a baseline, subjects were scanned with SQUID-based MEG machines at MGH. The aim is to make the exact same measurements using a nitrogen-vacancy sensor in a regular laboratory environment—that is, no cryogenics and minimal magnetic field shielding.

Making MEG as commonplace as other radiology techniques, such as X-rays, MRIs, and ultrasound, still requires significant research and development. Other potential alternatives to SQUIDs are atomic vapor cell magnetometers, which use a hot gas of alkali metal atoms to sense magnetic fields.

The ultimate vision, though still far off, would see MEG technology in widespread use as a diagnostic tool in hospitals, doctors' offices, and field medical facilities. The team continues to push their sensor's sensitivity and looks forward to further testing. If successful, nitrogen-vacancy diamond sensors could not only replace the current MEG technology in the scanners in the United States, but, Braje said, could also "put MEG measurements in hospitals and labs around the world."

A Nearly Neuron-Sized Brain Sensor

Lincoln Laboratory is developing what could be the world's smallest wireless, implantable device to measure neural activity in the brain

———

The brain's nearly 100 billion neurons, and the 100 trillion pathways connecting them, are responsible for telling the body precisely what to do, think, and feel. As such a vast and complex system, the brain leaves much for neuroscientists to still understand. However, moving fundamental research forward requires new technologies that enable a closer study of the brain. This need has been recognized by the U.S. National Institutes of Health, which have put nearly a billion dollars of funding toward their BRAIN (Brain Research through Advancing Innovative Neurotechnologies) Initiative since 2013.

Lincoln Laboratory researchers, too, have been pursuing the development of next-generation bioresearch tools. They are developing a new, tiny device—attempting to be as small as a single neuron—for studying neural activity. Implanted inside the brain, the device is designed to detect and wirelessly transmit the electrical signals it picks up from the neurons firing around it.

"One aspect of the overall effort in neuroscience to better understand the brain is about studying electrical signals in the brain, essentially asking, 'How do complex neuronal circuits work?'" said Farhan Adil, who leads this Microelectronics Interfacing
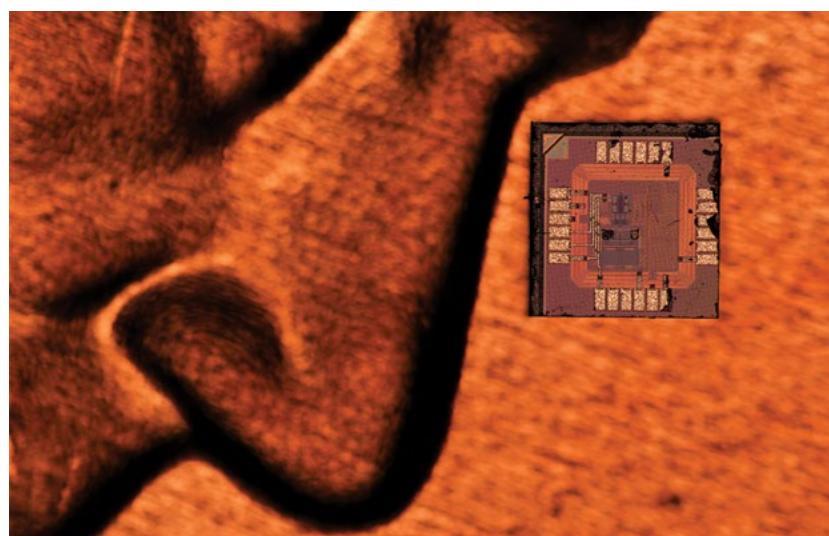
The application-specific integrated circuit, or ASIC, responsible for sensing voltage spikes in the brain is photographed here for scale on the face of a penny, next to Lincoln's nose.

Lab Notes

Neural Devices (MIND) program. "To answer this question, we need something very small to put into the brain, and it needs to have high spatial resolution and temporal resolution. So, we need a lot of them and they need to work fast."

Studying individual neurons in living mammals in real time and across the brain is challenging. The only current way to get high-resolution temporal and spatial data is through invasive means—removing part of the skull and implanting, for example, an array of electrodes directly into brain tissue. But these technologies are relatively large and often cause inflammation and scarring, disrupting the cells they are meant to observe. They also require bulky, wired connections to external receivers. "Our goal is to get the signal out wirelessly and truly make the device as passive and nonsurgical as possible," Adil said.

While not quite as small as the 0.1-millimeter size of a neuron, the current MIND prototype measures $1.5 \times 0.8 \times 0.4$ millimeters. If placed on the face of a penny for size comparison, the device would fit squarely on Abraham Lincoln's eye. At this size, the body's natural inflammatory response signal is unlikely to trigger, allowing the MIND device—or a handful of them, implanted around the brain— to operate essentially incognito to the neurons around it.

In simple terms, neurons communicate with each other by using electrical signals. A stimulus to the body is converted to an electrical signal, which rapidly travels down a neuron and is passed to other
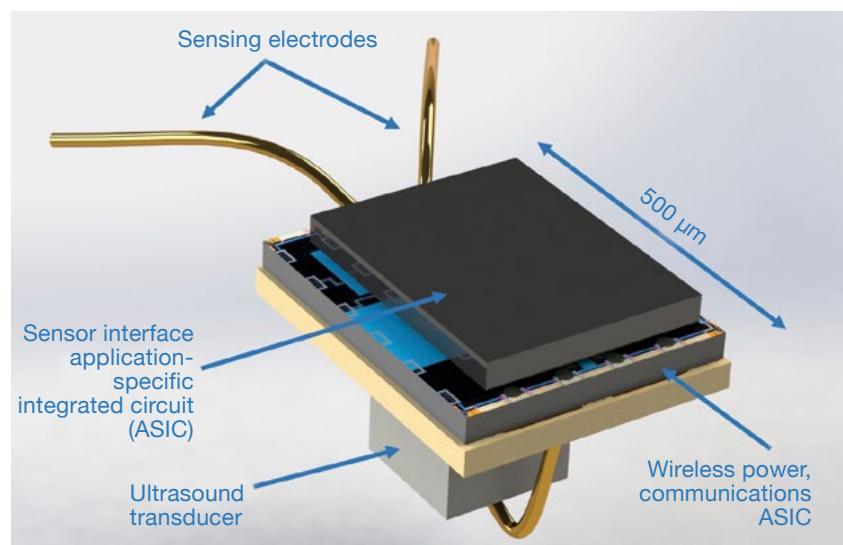
The researchers are working on building and testing a complete MIND prototype, with the sensor interface application-specific integrated circuit (ASIC) and sensing electrodes integrated with the wireless power/communications ASIC and the ultrasound transducer. Shown here is one vision of the final version. Not included in this illustration is the hydrogel that will coat the device.

neurons. This passing action, when the neuron "fires," creates a voltage spike. This spike is what the MIND device is designed to sense, capture electronically, and transmit out of the brain for researchers to study. To do so, the device relies on two custom application-specific integrated circuits (ASICs).

One of the ASICs is responsible for detecting signals and sits directly on top of the brain tissue. Electrodes from the ASIC sense the voltage spikes from firing neurons. When a voltage spike enters the ASIC, the tiny signal is amplified, converted from an analog to digital signal, and formally detected and time stamped. This sensor ASIC is bonded to a second ASIC that is responsible for relaying power to the sensor and for moving the electrical spike data out of the sensor.

For the ASICs to work, they need power. "The human head is

really good at keeping things out, like electromagnetic signals, for example," Adil said. "If you were to put a little RF antenna on the device, the head—with its hair, skin, skull, tissue—would greatly attenuate the signal." Instead, the team is using ultrasound frequencies to wirelessly power the device. These frequencies can penetrate the skull and brain with less signal loss than experienced by RF methods.

A transducer made out of piezoelectric crystal in the MIND device receives the incoming ultrasound frequency. The frequency causes vibrations in the transducer, generating an AC voltage that activates the power and communications ASIC. This ASIC converts the AC voltage to a DC voltage that powers the sensor ASIC, allowing it to collect the neural data.

Ultrasound energy is then used to transmit the data out of the

device by modulating, or varying the phase of, the ultrasound waves that bounce off the MIND transducer. "Not all of the ultrasound energy is collected by the piezoelectric crystal. Some of it will be backscattered," said Mark Hernandez, who has been supporting the program by testing the device components. "We hope to encode the data, 0s and 1s to indicate whenever there is an electrical spike, into that backscattered signal." This signal would be collected again by an ultrasound receiver outside the body and decoded.

If placed on the face of a penny for size comparison, the device would fit squarely on Abraham Lincoln's eye.

The team is currently demonstrating that they can create sufficient electrical power via ultrasound energy for a device this small. Their modeling shows their device works. "But transmitting and transducing enough ultrasound energy to power the sensor circuitry is very difficult," Hernandez added. The system must have a very low power draw of less than 10 microwatts. If the device is overpowered, it could generate too much heat to be biologically safe. Also, the Food and Drug Administration sets limits on the amount of ultrasound energy to which a person can be exposed.

Once the power demonstrations are complete, the team will turn their attention to miniaturizing the transducer even more (it is currently the largest component of the device). "Our hope is to design a new type of ultrasonic receiver that can be microfabricated and integrated with a conventional complementary metal-oxide semiconductor electronics process," said Siddhartha Ghosh, who is leading the transducer efforts.

Meanwhile, the sensor ASIC has been tested successfully in the lab. Hernandez and colleague Christina Zook led the experiments, first using cultured rat cardiomyocytes, cells that make up the heart muscle and that give off high levels of voltage when electrically stimulated. "We could see that the sensor was picking up the signals from the cells. It could even measure a change in activity. For example, we added norepinephrine, which increases the 'beating' rate of the cells," Hernandez said. The sensor was also able to detect the low-voltage firing of active neurons in subsequent in vitro tests. The results matched those of a gold-standard microelectrode array.

While there are still engineering challenges to overcome, Adil hopes to soon transition a complete MIND prototype out of Lincoln Laboratory and into the research space. In giving neuroscientists a new tool to understand the neuronal circuitry of the brain, the device could also help unveil new connections between brain function and neurological disease. "This new tiny, wireless tool could really help push fundamental neuroscience research to the next level," he said. "There's not a wireless neural sensor of this size out there at the moment."

WEARABLE SENSORS

Progress Toward Wearable Pulse Oximetry

New algorithms and functional fibers clear a way for monitoring blood oxygen saturation on the go

---

During a routine medical visit, you will likely encounter a pulse oximeter. The sensor is clipped onto your finger to measure heart rate and the saturation of oxygen in your blood, two indicators of cardiorespiratory health. Pulse oximeters are a standard tool for monitoring people in a hospital. However, taking these measurements while a person is on the move, out in the world, remains a challenge.

"Try moving your hand around with the sensor on, and the doctor will ask you to sit still," said Brian Telfer, a senior staff member in the Human Health and Performance Systems Group. Motion corrupts the signal recorded by the oximeter and is one of the main reasons why pulse oximeters have yet to be turned into a wearable technology.

Yet, a wearable pulse oximeter is something both the military and civilian populations need. Real-time oximetry could help predict if

soldiers or hikers are experiencing altitude sickness and need medical attention, or if pilots are suffering from hypoxia and need help before an accident occurs. There are trauma-related applications as well. Changes in the raw oximetry waveforms may be used to detect internal hemorrhaging, which is difficult to anticipate from standard vital signs.

To address these needs, Telfer and his colleagues are teaming with the U.S. Army Research Institute of Environmental Medicine to prototype a wearable pulse oximeter. They are tackling the problem of motion by developing algorithms that can determine the difference between frequencies that are caused by motion and those that reflect heart rate and blood oxygen saturation, referred to as SpO2.

"The crux of our work is to have an accelerometer co-located with the sensor," said James Williamson, who is developing these algorithms. "The accelerometer senses motion and at what frequency. If we can locate and get rid of that frequency in the sensor's data, then we can determine heart rate and SpO2 from the signal."

The signals Williamson refers to are optical signals, the same kind used by commercial wearable devices such as Fitbits to estimate heart rate. A light-emitting diode, or LED, reflects onto the wearer's skin, and a photodetector captures changes in light absorption as blood pulses underneath the skin. Algorithms use these changes in light absorption to calculate heart rate (though studies show that these commercial systems' accuracy also
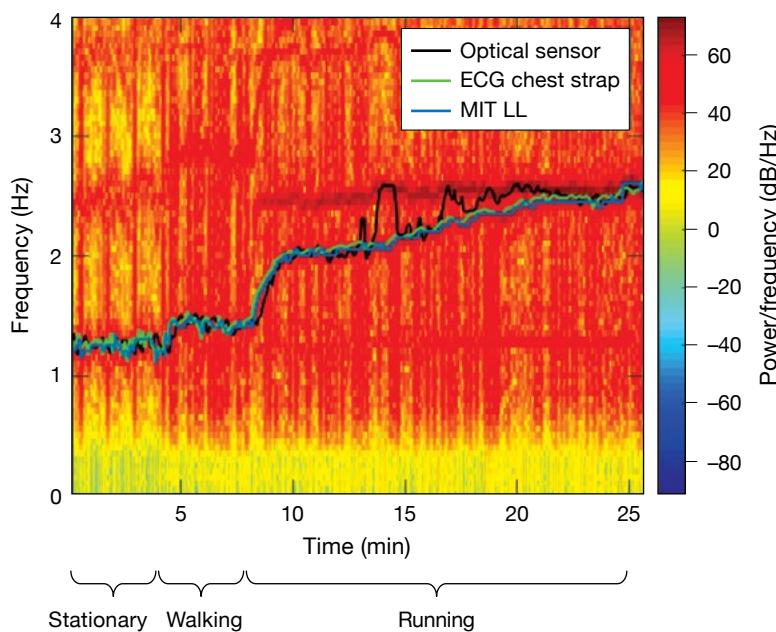
A spectrogram of the infrared channel from a commercial pulse oximeter chip worn at the forehead shows frequency components from both the wearer's heart rate and from motion artifacts while running. Overlaid are heart rate frequencies measured from the same data collection from a commercial optical sensor (black), an electrocardiogram chest strap (green), and the pulse oximeter chip with Lincoln Laboratory's motion mitigation system applied (blue). The chest strap and the Lincoln Laboratory system accurately follow the heart rate component, while the commercial optical sensor sometimes confuses the heart rate and running components.

suffers from motion artifacts). A similar process is used to measure SpO2, though the device needs to use both red and infrared light, which oxygenated and deoxygenated blood absorb differently.

Separate algorithms for determining heart rate and SpO2 during motion were developed. To assess their algorithms' performance, the researchers had test participants wear headbands or helmets outfitted with a three-axis accelerometer and a commercial pulse oximeter chip pressed against their foreheads. Data were collected as a person stood, walked, and ran on a treadmill. For comparison, the researchers also had each participant wear a hospital-grade pulse

oximeter clipped to a finger and an electrocardiogram chest strap, the gold standard for obtaining heart-rate measurements.

As was expected, the chest strap sensor correctly followed the heart-rate frequency, and the finger-worn optical sensor jumped between following the heart-rate frequency and the movement frequency. But when the Laboratory's algorithm was applied to the data recorded by the forehead chip, it allowed the sensor to consistently follow the participant's heart rate and ignore motion-caused frequencies. "We're showing with our method that we can actually measure heart rate better than optical-based commercial systems," Williamson said.
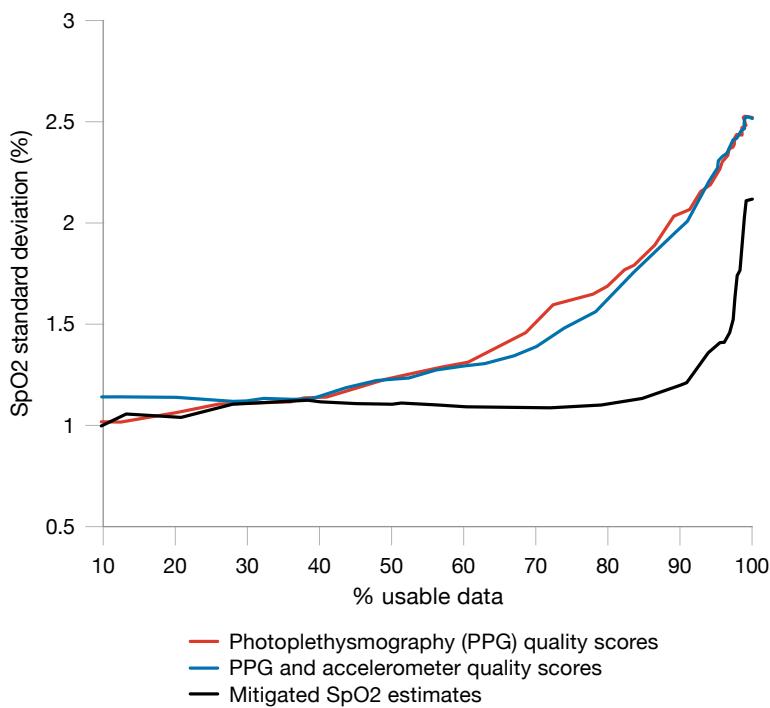
Lincoln Laboratory's mitigation techniques allow for more of the sensor data captured in the presence of motion to be used to accurately estimate SpO2. For normal physiology, SpO2 changes slowly over time, so the standard deviation of SpO2 over a short time window can be used as a metric of SpO2 accuracy. With the mitigation technique applied, the SpO2 standard deviation remains low even when 90 percent of the data is used. In contrast, the SpO2 standard deviation is significantly higher when using alternative techniques, such as optical (photoplethysmography [PPG]) quality scores and PPG and accelerometer quality scores.

A related algorithm was developed to mitigate motion artifacts in the SpO2 computation. The algorithm looks at patterns in 5-second windows of data, analyzing the repetitive peaks in the infrared and red optical frequencies. It then compares these peaks to those in the accelerometer frequencies. If the peaks are too similar, that frame's data are deemed corrupted and are thrown out. The algorithm instead uses nearby uncorrupted 5-second frames to compute SpO2.

SpO2 is calculated as the percentage of oxygenated hemoglobin in the blood; above at least 95 percent is normal. After test subjects began running, average SpO2 estimates from the commercial finger-worn oximeter dropped from an average 97 percent to 93 percent, likely indicating motion corruption. With the Laboratory algorithm applied to the forehead-chip data, SpO2 estimates for test subjects remained consistently at or near 98 percent.

"We are doing more testing to validate the SpO2 measurements," Williamson said, "but the motion mitigation results are promising for enabling wearable pulse oximetry."

The team has also begun turning their attention to

prototyping a wearable pulse-oximetry system. Instead of a traditional chip-based device, the system will take advantage of a new area of research and development at Lincoln Laboratory: advanced functional fibers. Here, they are partnering with the Laboratory's Defense Fabric Discovery Center (DFDC) to weave advanced, optical-sensing fibers into the very fabric of a soldier's uniform.

"Many chip devices, like the ones worn on a wrist, are rigid," said Lauren Cantley, a materials scientist who works in the center. "We want a comfortable system and one that can mate to the surface of your skin."

Certain parts of the body are better than others for capturing heart rate and especially SpO2 measurements; a fabric-based system allows the team to expand the possibilities of sensor locations. The fabric's flexibility also helps to ensure the sensing elements are in constant contact with the skin, improving signal quality. Initial investigations by the team have found that the forehead is an ideal sensing location because the reflectivity of the bone yields high-quality signals (thus the team's decision to test their algorithms with a chip pressed to the forehead) and reduces power needs. But once the pulse-oximetry textile is fabricated, more testing can be done on various areas of the body.

To make this textile, DFDC staff are building on a breakthrough technology developed in conjunction with MIT and the Advanced Functional Fabrics of America institute. They embed semiconducting

devices—LEDs and photodetectors—into a block of polycarbonate plastic that will be drawn into a fiber. The tiny devices are placed along hollow channels drilled through the block, then copper wire is fed through these channels. As the block is heated and pulled into a thin fiber, the devices are forced into electrical contact with the wire. This fiber now has semiconducting capabilities throughout its length.

To date, the team has proven this fiber's use for optically measuring heart rate. When a user places a finger over a fabric woven with these fibers, the light emitted from the LED-embedded fiber is partially absorbed by blood and reflects back up to a photodetecting fiber in the fabric. Current efforts are focusing on demonstrating that SpO2 can be measured using this fiber technology.

Integrating a pulse oximetry capability into a uniform also has the benefit of not requiring a soldier to carry an additional sensor. "We're adding functionality to something they already have on their body," Cantley said. Along with pulse oximetry, the integration of other physiological sensors into fibers and then into a soldier's uniform is something the DFDC staff are actively working on. The goal is to build a full "body area network," a system of sensors to continuously transmit physiological status updates to medical staff without adding burdensome equipment to servicemembers in the field.

Telfer sees a wearable pulse oximeter being commercially successful as well. The surge in popularity of wearable devices points to people's desire to have real-time data about their health. But it's not just for the fun of knowing—these data can save lives. "Motion mitigation is just one type of the signal processing we are doing," Telfer said. "For blood loss, for example, the Army and other researchers have found that the shape of the waveform changes." Processing these waveform changes and alerting a user may mean that vital steps can be taken before blood loss leads to shock. The researchers are looking into other signal processing techniques that could give greater insight into a user's health.

If a pulse-oximetry textile is realized, it is possible that the clothes you wear to a future doctor's visit will have already told you what the finger clip reveals. Until then, sit still.

Fibers drawn from a preform with embedded LEDs are woven into fabric by using a conventional weave.

DATA ANALYSIS

Expanding the Potential of Medical Data with Polystores

A new approach to integrating medical information from multiple databases aims to improve diagnoses and treatment

Big data, the explosion of digital data available today, has led to significant challenges caused by what is termed the three Vs: volume, velocity, and variety. The massive volume of data generated hampers a system's ability to process, index, and store it all. Datasets streaming into computer networks at high velocity hamstring processing, and the now

commonplace diversity of data—text, images, graphs, and video, for example—constrains a system's capability to manage files efficiently.

No domain is more challenged by these factors than the medical community whose systems must integrate and synthesize a rapidly expanding collection of data to support the decisions medical personnel make. Sensor data, such as MRI scans or electrocardiographs (ECGs), physicians' reports, patient histories, and research findings are just some of the data that are readily available in digital formats.

To deal with this variety of data, programmers working for the medical establishment have developed custom databases for the sensor data, graphs, documents, and transactions that are recorded in medical files. These custom databases deliver 100 times better performance than can general-purpose databases, but the trade-off is that these performance benefits have led to a profusion of data-specific storage engines.

"Most modern decision support systems contain five or more distinct data storage systems," said Vijay Gadepally, a senior member of the technical staff in the Lincoln Laboratory Supercomputing Center (LLSC). Large medical establishments, particularly regionalized ones, may employ even more than five. In addition, for organizational or policy reasons, data are often required to be housed in different databases. "For an application developer, this situation translates to developing system-specific interfaces and connectors for every different system," Gadepally

BigDAWG signals a big shift in the way we think about data management. It simplifies the integration of data distributed across multiple database engines by leveraging translators across common programming and data models.

explained. Moreover, developing the applications is complicated by the developers' need to understand the nuances of each underlying system.

Gadepally and colleagues in the LLSC were part of a consortium of researchers from Lincoln Laboratory, MIT, and several other universities that has been developing a system composed of multiple storage engines able to support queries spanning multiple data models. Called a polystore, this system is proving to be a viable alternative to the more traditional parallel and federated database systems.

The system developed in this collaboration, BigDAWG (short for the Big Data Working Group), is a reference implementation of a system designed to simplify database management for complex applications. BigDAWG is predicated on several concepts about a solution to the database management problem: one database will not fit all; no single programming language suits all applications; integrating multiple storage engines should not impede a system's functionality; and applications should be portable across different systems. The research team, which looked at using the system for several applications,

aimed BigDAWG at the challenge to decision making posed by the increasingly large and diverse amount of data available to the medical community.

They explored how BigDAWG could work with the Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC II) medical dataset. MIMIC II contains physiological signals and vital signs captured from patient monitors of tens of thousands of patients in intensive care units. MIMIC II also contains patients' metadata (e.g., name, age), the text of doctors' and nurses' notes, lab results, and information about patients' prescriptions. Because a hospital typically stores all the patients' historical data and real-time sensor feeds, the hospital's data management system is taxed to support several databases and types of data processing: standard SQL analytics (e.g., how many patients were given a particular drug), complex analytics (e.g., compute a patient's ECG data and compare the results to "normal" ECG readings), text search (e.g., find patients' responses to a drug), and real-time monitoring of physiological conditions.

Each of the underlying databases hosts part of the MIMIC

II data corpus—epitomizing "one size does not fit all"— and MIMIC II's multiple databases present a good example of how no one database can comprehensively support complex decision making.

The team applied the polystore solution to a medical analysis that is looking for patients with similar physiological measurements so that doctors might find data useful for predicting which patients may be likely to experience a deterioration of the circulatory system. Using BigDAWG, the analytic first performed a discrete wavelet transformation of an individual patient's electrocardiogram record; then the system generated a wavelet coefficient histogram, performed a function that weights rare ECG changes higher than those commonly seen in ECGs, and correlated these changes across all other patients. Each of these subanalytic pieces was then split across different systems, and the performance of the system was measured.

For the MIMIC II demonstration, the team built a number of interfaces. The browser let the user scan through the entire MIMIC II dataset or drill down to more detailed information. A basic analytics tool enabled the user to spot and explore relationships in medical data, while a tool for complex analytics allowed users to perform data correlations useful for advanced research. Interfaces that facilitate keyword searches and real-time monitoring of sensor data were also included.

"Our results support the notion that overall query performance may be improved by identifying
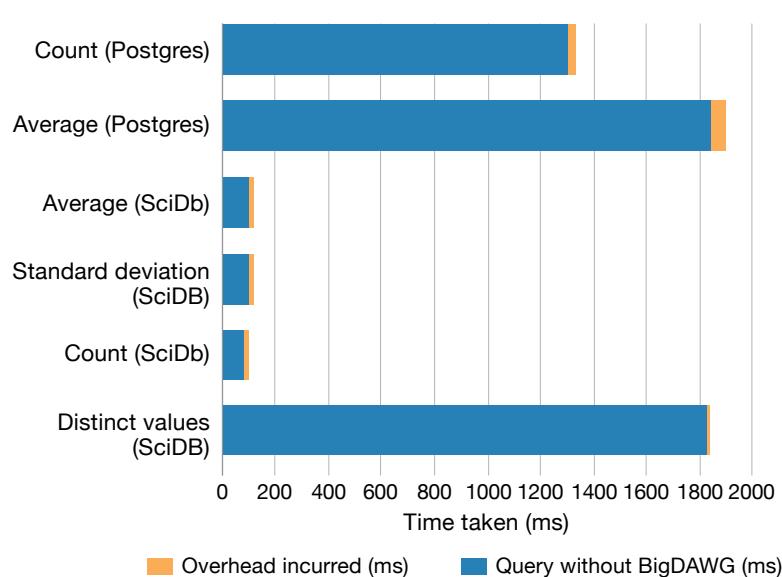
(a)
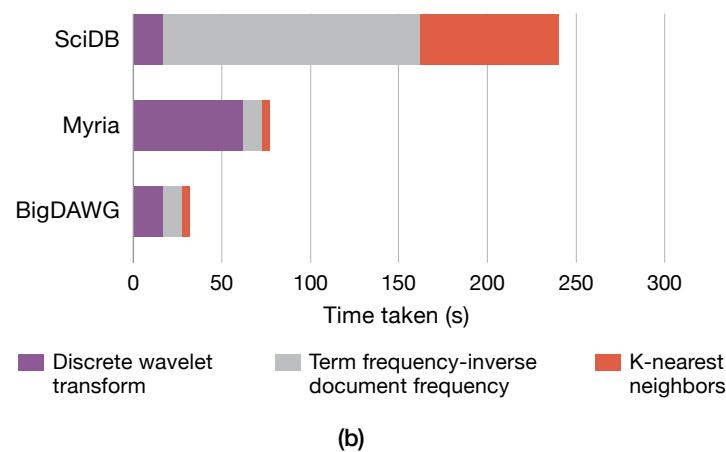
(b)

The charts depict the BigDAWG performance on MIMIC II analytics. The BigDAWG computational overhead for common database queries is presented in (a). The blue bar indicates the time taken without BigDAWG and the orange bar indicates the additional time incurred when using BigDAWG. A comparison of BigDAWG performance for a medical analytic against state-of-the-art SQL and NewSQL databases is seen in (b). For this particular analytic, BigDAWG can cut the overall compute time significantly. V. Gadepally et al., "The BigDAWG Polystore System and Architecture," *2016 IEEE High Performance Extreme Computing Conference*.

and leveraging relative strengths of disparate database systems within a polystore, and that this improvement does not significantly decrease performance," Gadepally said.

BigDAWG automatically optimizes cross-engine queries

and moves data as needed across disparate engines without user intervention. "BigDAWG signals a big shift in the way we think about data management. It simplifies the integration of data distributed across multiple database engines

by leveraging translators across common programming and data models," Gadepally said.

BigDAWG users can write their applications in popular programming languages for databases, such as SQL, Graph, Array, and Text, because BigDAWG's tools can translate queries among different systems. Although BigDAWG integrates data from diverse engines, it allows developers to exploit the full set of capabilities and the full performance of any database underlying their applications. This benefit is achieved because the BigDAWG modular architecture permits users to customize their database usage; a middleware layer handles various queries, and "islands"—which comprise a data model, a query language or set of operators, and one or more database engines—confine the programming to just the portions of the multiple databases necessary to enable the application.

"We believe that polystores such as BigDAWG will play a big part in the future of complex data management," Gadepally said. "We've held successful workshops at the 2016 and 2017 IEEE International Conferences on Big Data and the 2018 and 2019 Very Large Data Bases conferences. Through these and other outreach efforts, we have been surprised by the number of disciplines and research efforts that are looking for complex data management solutions," he added.

For the near future, the BigDAWG researchers are investigating applications of relevance to the U.S. Navy, Department of

Polystores such as BigDAWG will play a big part in the future of complex data management.

Veterans Affairs, and Air Force, and are looking at ways to introduce security and access control for polystore systems.

For the medical community, a polystore solution like BigDAWG, which is available as open-source software at https://bigdawg.mit. edu, may allow developers to create applications that provide clinicians with fast analytics to aid in making decisions about diagnoses and treatments, and that enable medical researchers to make new discoveries.

Noninvasive Biomarkers of Neurobehavioral Performance

Thomas F. Quatieri, James R. Williamson, Adam C. Lammert, Kristin J. Heaton,

and Jeffrey S. Palmer

A large array of neurological and psychological conditions is prevalent in civilian and military populations. To complement current clinical standards, there is a pressing need for noninvasive in-field and at-home methods of assessing such conditions. Lincoln Laboratory is developing neurobehavior-based biomarkers, which reflect a change in brain functioning as manifested in motor control, paired with neurocomputational biophysical models to identify neurobehavioral changes. The efficacy of our approach is illustrated through several applications for assessing major depressive disorder, Parkinson's disease, traumatic brain injury, and cognitive overload.

Mental health impairments can significantly and adversely affect human performance and quality of life for civilian and military populations. Indeed, brain disorders taken together are a leading cause of global disease [1]. Toward the goal of finding simple, noninvasive, and objective means to detect, predict, and/or monitor such conditions, the Human Health and Performance Systems Group at Lincoln Laboratory is developing multimodal biomarkers based on behavioral measurements to detect changes in neurobehavioral function associated with psychological (e.g., major depressive disorder), neurotraumatic (e.g., traumatic brain injury), neurodegenerative (e.g., Parkinson's disease), and neurodevelopmental (e.g., autism spectrum disorder) conditions.

One of our objectives is the discovery of behavioral-based biomarkers that reflect a change or decline in brain functioning as manifested in motor control and, more specifically, changes in timing and coordination within the neuromotor components of a behavior. This investigation into biomarkers is based on the hypothesis that motor control is associated with neural coordination across different parts of the brain. Examples of behaviors we are examining include vocal and facial expression, heart rate variability, eye movement, and fine and gross movement of the extremities. One focus of our investigation is on novel vocal and facial biomarkers that are based on phonetic timing, articulatory coordination, and facial muscle coordination during speech production.

A motivation for working in the areas of health and performance is the prevalence in civilian and military

THOMAS F. QUATIERI, JAMES R. WILLIAMSON, ADAM C. LAMMERT, KRISTIN J. HEATON, AND JEFFREY S. PALMER

populations of conditions that can compromise neuro-behavioral function. Among the neurobehavioral disorders that affect global health, depression is the single largest source of lost productivity in high-income countries. According to the World Health Organization [1], by the year 2030, disability and lives lost from depression alone will surpass those caused by cancer, war, stroke, and accidents. There are roughly 20 suicides per day among U.S. veterans [2]. It also is projected that by 2030 there will be 82 million people worldwide with dementia [3]. One in 59 children in the United States has been identified as having autism spectrum disorder (ASD) [4]. About 19 percent of U.S. servicepeople returning from recent wars experienced a traumatic brain injury (TBI) of some form, and about 20 percent of that same population has post-traumatic stress disorder (PTSD) [5]. Other factors, such as (1) exposures to environmental extremes or occupational chemicals and (2) fatigue resulting from physical and mental exertion or disrupted sleep patterns, can also compromise neurobehavioral performance in our service-people. Figure 1 illustrates the staggering and increasing prevalence of mental health conditions and other factors that can impair neurobehavioral performance.

The many standard methods used in detecting neurobehavioral performance changes range from brain imaging to clinical assessments to molecular diagnostics. These approaches, while useful, can be time and resource intensive, are often susceptible to the effects of individual motivation, and, because they often capture feelings or behaviors at discrete points in time, may lack sufficient sensitivity to detect subtle changes in behavior. Moreover, they often lack objective measurement, especially in early detection of day-to-day performance changes when assessment can be most important. At Lincoln Laboratory, we are developing technologies that address these limitations by reaching large populations and detecting, monitoring, and ultimately intervening to follow the effect of treatment and intervention. Our approach seeks objective biomarkers that reflect subtle changes in behavior and makes use of nonobtrusive mobile wearable technologies.

In the context of this article, the term biomarker refers to any measurement of behavior we can obtain from a human body, such as talking or walking. Certain behavioral features have been shown to change with a subject's mental and emotional state and under numerous conditions, including cognitive load and neurological

5.8 million
U.S. adults are living with Alzheimer's. By 2050, this number is projected to rise to nearly 14 million.

7.6%
of Americans ages 12 and older had depression during 2009–2012.

1 in 59
eight-year-old children in 11 communities across the United States were identified as having autism in 2014.

8 million
U.S. adults have post-traumatic stress disorder (PTSD) during a given year. This is only a small portion of those who have gone through a trauma.

50,000
people are diagnosed with Parkinson's disease in the United States each year and about half a million people have the disease.

19.1%
of U.S. adults had an anxiety disorder in the past year.

19.5%
of U.S. service members who have returned from Afghanistan and Iraq report experiencing a traumatic brain injury during deployment.

Sources: Alzheimer's Association, "2019 Alzheimer's Disease Facts and Figures." National Institute of Mental Health, "Any Anxiety Disorder," available at https://www.nimh.nih.gov/health/statistics/any-anxiety-disorder.shtml, accessed Sept. 2019. J. Baio, L. Wiggins, D.L. Christensen, Z. Warren, et al., "Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014," Morbidity and Mortality Surveillance Summaries, vol. 67, no. 6, 2018, pp. 1–23. National Institutes of Health, "Parkinson's Disease Fact Sheet," available at https://report.nih.gov/NIHfactsheets/ViewFactSheet.aspx?csid=109, accessed Sept. 2019. L.A. Pratt and D.J. Brody, "Depression in the U.S. Household Population, 2009–2012," National Center for Health Statistics Data Brief No. 172, 2014. U.S. Department of Veterans Affairs, "How Common is PTSD in Adults?" available at https://www.ptsd.va.gov/understand/common/common_adults.asp, accessed Sept. 2019.

FIGURE 1. An array of neurological and psychological conditions is prevalent in U.S. civilian and military populations.

disorders. Example modalities used in detecting cognitive and neurological stress include voice, facial expression, physiology, eye movement, gait, hand dexterity, and electroencephalography (EEG) analysis. The use of behavioral-based biomarkers is not a new concept. Such biomarkers have been applied in a variety of contexts for at least two decades. Examples include features derived from vocal and facial expression, fine and gross motor movements, physiology, and brain activity [6–40].

Features derived from vocal expression, or voice, include characterizations of prosody (e.g., fundamental frequency and speaking rate), spectral representations (e.g., vocal tract resonance), and glottal excitation flow patterns (flow shape, timing jitter, amplitude shimmer, and aspiration) [10, 13, 14, 33, 36] derived from acoustic measurements. Features from facial expression include parametric models of the face and representations of underlying facial muscle groups known as facial action units [22, 24] derived from video measurements. For fine and gross motor movements, irregularities in stride, hand dexterity, and eye tracking have been used in characterizing a variety of neurological conditions, with a few examples in Russo et al., Samadani et al., and Gowan and Hamilton [40–42].

We cannot hope in the limits of this article to review all state-of-the-art behavioral-based feature approaches. However, what distinguishes Lincoln Laboratory's approach from these more standard methods, or what we will refer to as low-level feature-based approaches, is the introduction of features motivated by the timing and coordination of underlying neuromotor control of behaviors. Although significant effort has been devoted to behavioral-based biomarkers, little or no study has been done examining changes in coordination, movement, and timing of components of a behavior. For example, in individuals suffering from depression, neurophysiological changes often alter motor control and thus affect mechanisms controlling speech production and facial expression. Clinically, these changes are typically associated with psychomotor slowing, a condition of slowed neuromotor output manifested in altered timing and coordination across multiple observables of acoustics and facial movements during speech. We refer to features based on this paradigm as high-level features.

While we begin in this article with standard low-level features in each modality, we build upon these using high-level timing and coordination features that reflect underlying neural activity across the brain. It is hypothesized that these relations are associated with neural coordination across different parts of the brain that are essential in motor control. Various subsets of these features have been used effectively at Lincoln Laboratory for detecting neurobehavioral changes associated with depression, Parkinson's disease, TBI, and dementia [13, 43–46], as well as mental exertion under stress [43, 47], and thus perhaps form a common feature basis for neurocognitive change.

Neurophysiological Basis and Framework

Our approach involves biomarkers of human behavior that we can observe from the human body. At the Laboratory, we are investigating a number of different behaviors, broadly defined to include fine motor movements (e.g., hand and finger dexterity and eye tracking), gross motor movements (e.g., balance and gait), skin conductance, and heart rate. But we are focusing in this article on vocal and facial expression, and specifically on markers that reflect change in motor aspects of brain function.

Neuromotor Representations in the Brain

In our approach, we seek biomarkers that satisfy two primary properties:
1. They reflect decline in brain functioning as manifested in motor control measured from bodily behaviors.
2. They reflect changes in timing and coordination both within and across components of a behavior.

Deriving biomarkers from vocal and facial expression is desirable for a number of reasons: vocal and facial expressions are easily measured, noninvasive, and accessible, and, importantly for our approach, they are highly complex human behaviors that require precise coordination across different regions of the brain. In speaking, for example, the articulators (tongue, lips, jaw, and velum) are finely coordinated, and this coordination can change under conditions of injury, illness, and stress. Underlying this articulatory complexity is the even more complex control system of the brain [48]. As an example, we overview this control system in speech production. There is evidence of a similar kind of neural network complexity that controls facial expression during speaking, as well as in general expressiveness in paralinguistic socioemotional communication [49].

THOMAS F. QUATIERI, JAMES R. WILLIAMSON, ADAM C. LAMMERT, KRISTIN J. HEATON, AND JEFFREY S. PALMER

As schematized in Figure 2, the core speech-production network consists of many components, widely distributed throughout the brain. These components go from developing a concept and selecting sentences and words to express a concept, to deciding syllables and phonemes to represent words, to positioning and coordinating articulators, to firing neural signals that activate muscles to move the articulators. Finally, there are auditory and somatosensory feedback mechanisms used to monitor and self-correct speech production. Somatosensory includes proprioception (a person's sense of position and movement of the body) and tactile feedback (a person's sense of tongue placement and vocal cord vibration).

In addition to this core production network, nonspeech regions of the brain modulate the core. Given this wide distribution of modules in the brain that either directly or indirectly control speech production, it is likely that one of the modules will be affected by neurological or stress conditions, and speech production will suffer. In fact, all of the brain disorders mentioned at the start of this article cause some kind of speech degradation. For depression, as an example, there is evidence that a nonspeech region called the limbic system, which controls mood and emotion, is disrupted. This disruption propagates to the core speech network, in particular to neural circuits that control timing and coordination of the articulators. The general observation by clinicians that depressed individuals, on average, tend to talk slower and have less clarity in their articulation than do nondepressed individuals is consistent with a disruption of timing and coordination circuits. Motivated by these observations, we have introduced two novel biomarkers that are based on the decline in neuromotor timing and
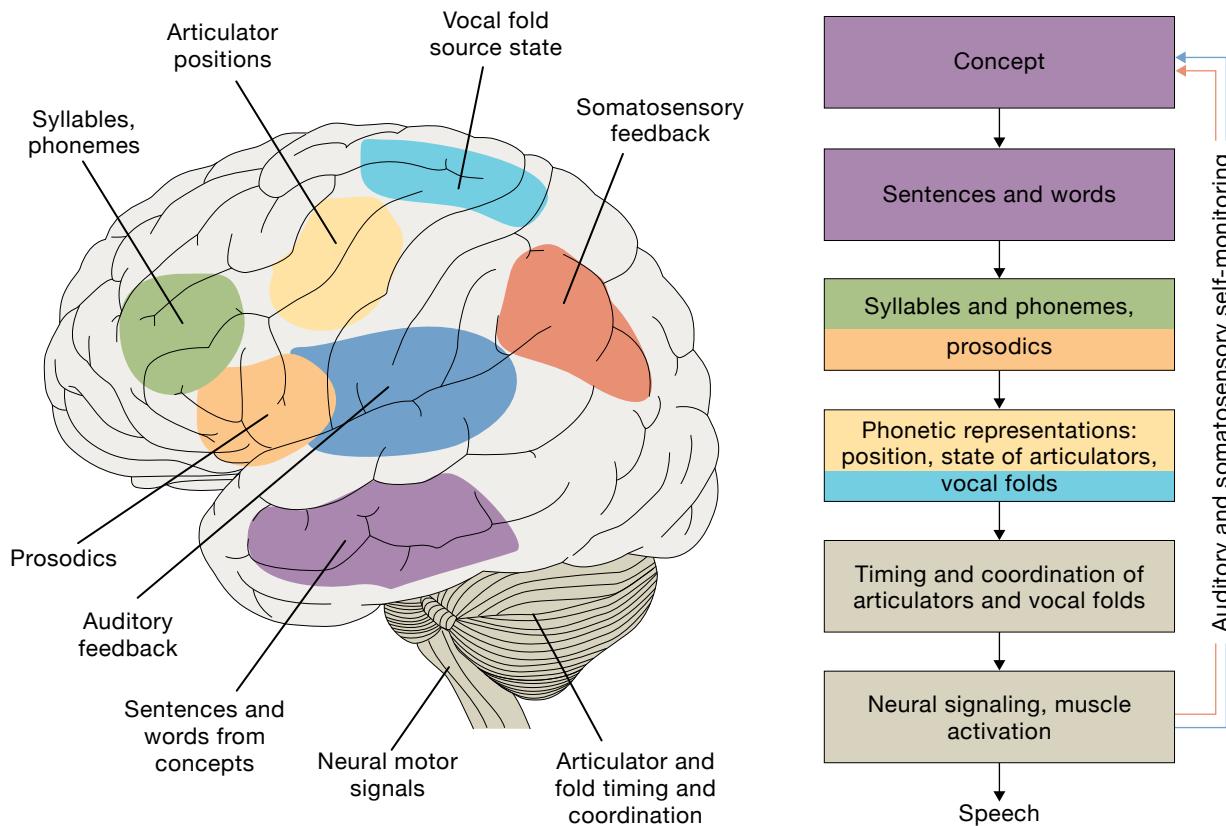
FIGURE 2. The figure shows a simplified view of the core speech production neurocognitive network with many components widely distributed throughout the brain. The somatosensory region is associated with tactile and proprioceptive (a sense of position and movement of the articulators) sensory feedback. Nonspeech regions, such as visual and cognitive areas (not shown), can modulate the core speech network. It is important to note that Figure 2 is a gross oversimplification of the speech brain network that involves complex coordination across multiple brain regions.

coordination that we will describe. We also consider these biomarkers more generally across all conditions.

Framework

The new features and the corresponding detection system that we have designed leverage many years of Laboratory expertise in speech and facial signal processing and automatic classification. At a bird's eye view (Figure 3), our detection system first estimates a set of standard vocal and facial features. These features that leverage our and others' past work [50–52] we call low-level features. Examples of low-level features from an acoustic speech signal are vocal tract resonances (termed formants in the speech community) and automatically derived speech phoneme labels (units of sound that distinguish one word from another). Low-level features from facial video are facial action units that reflect muscle groups associated with basic elements of expression (lowered brow, puckered lips). From these low-level features, we then extract our new features on the basis of a timing and coordination concept, and we call these high-level features. With the high-level features, we train a classifier that provides us with a binary decision or severity-level estimate of a condition.

Feature Extraction
Vocal Features

The categorization of vocal characteristics is broken down into three components: speech source (at the vocal folds), system (vocal tract), and prosody (sometimes referred to as the melody of speech). As shown in Figure 4, the lungs provide the airflow that assists in making the vocal folds vibrate, sending a periodic or noise-like stimulus to the vocal tract. The vocal tract—consisting of the oral, nasal, and pharynx subsystems—provides "color" to the sound, with different tract shapes yielding different phonemes. Prosody is a function of pitch, timing, and energy fluctuations.

We exploit dynamic variation and interrelationships across speech production systems by computing features that reflect complementary aspects of the speech vocal-fold source, vocal tract system, and prosody [50]. Across the three categories, we use a broad suite of features that are used in the detection of neurological disorders and neurobehavioral performance changes associated with a wide range of exposures and other sources of stress [13, 43–47, 53–60]. In each category,
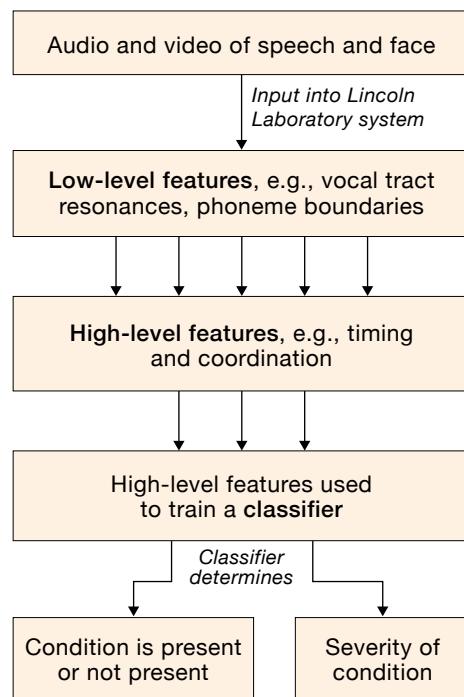
FIGURE 3. Lincoln Laboratory's system can detect the presence or estimate the severity of a neurobehavioral condition. Low-level features are first derived from audio and video of speech and face, respectively. Next, high-level features are extracted from neuromotor timing and coordination features. These high-level features are then used to train a classifier to determine the presence and/or severity of a neurobehavioral condition.

we leverage low-level and high-level feature types that were introduced previously.

LOW-LEVEL FEATURE EXTRACTION

The various low-level features are estimated from windowed speech segments frame-by-frame over time, and the window type (e.g., Hamming, Hanning, Kaiser), length (typically between 10–40 ms), and sliding frame (typically between 1–10 ms) are selected depending on the feature type.

Many standard low-level features used in the speech community characterize the degree of periodicity of the vocal-fold vibration within the vocal source. One such feature is the harmonics-to-noise (HNR) ratio, which is the ratio, in decibels (dB), of the power of the harmonic (periodic) signal from vocal-fold vibration and the power of the speech noise signal at the vocal folds created by turbulence as air rushes past the vocal
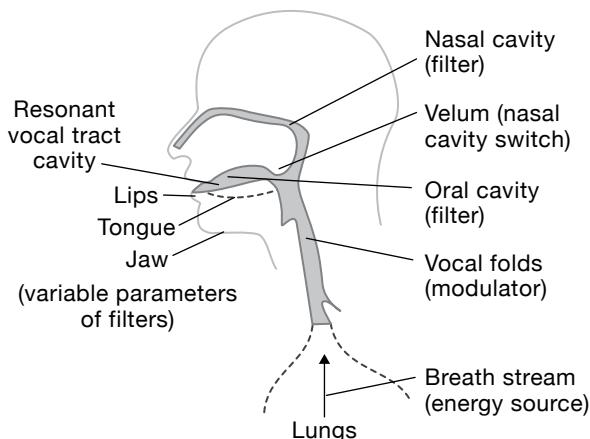
FIGURE 4. This illustration shows the speech source (at the vocal folds) and system (vocal tract) [50]. Prosody is a function of pitch, timing, and energy fluctuations. The lungs provide an energy source and the airflow that assists in making the vocal folds vibrate, providing input to the vocal tract. The vocal tract—consisting of the oral, nasal, and pharynx subsystems—provides "color" to the sound created at the vocal folds, with different tract shapes yielding different phonemes.

folds from the lungs. This measure is thought to reflect breathiness in a voice. Our HNR is computed by using a periodic/noise decomposition method that employs a comb filter to extract the harmonic component of a signal [61, 62]. Two other low-level source features are (1) cepstral peak prominence (CPP), which is defined as the difference in dB between the magnitude of the highest peak and the noise floor in the power cepstrum (the result of taking the inverse Fourier transform of the logarithm of the estimated spectrum of a signal) and (2) a measure called creak, corresponding to what is often referred to as creaky voice, which reflects large irregularity in pitch periods (often with low average pitch) and high peakiness of airflow pulses that excite the vocal tract [63, 64].

Low-level vocal tract–based features are designed to reflect the intensity and temporal dynamics of the vocal tract frequency response. One primary feature set comprises the vocal tract resonances (referred to as formant frequencies) estimated by a Kalman filter technique, smoothly tracking the first three formant frequencies while also smoothly coasting through nonspeech regions [28]. A second primary feature set is referred to as Mel-frequency cepstral coefficients (MFCCs) that provide frequency response intensity, while

16 delta MFCCs [50] are used to characterize velocities of vocal tract spectral magnitudes [65]. Delta MFCCs are computed by using regression with the two frames before and after a given frame.

The final low-level feature types reflect speech prosody. The first feature type is pitch (also referred to as fundamental frequency), which we estimate by using a time-domain autocorrelation method [50]. Our second prosodic-based feature type relies on an automatic phoneme recognition algorithm [66]. This algorithm obtains phonetic boundaries and phoneme labels with one of 40 phonetic speech classes detected and allows us to obtain average speaking rate (number of phonemes per second).

HIGH-LEVEL FEATURE EXTRACTION

Our high-level features, often derived from low-level features, are designed to capture timing, coordination, and fine time resolution of dynamics of speech production components. We refer to one high-level feature type as correlation structure, which is a function of the temporal correlation (on different time scales), reflecting a form of coordination within and across vocal source, system, and prosodic speech components (illustrated in Figure 5).

In this approach, channel-delay correlation and covariance matrices are computed from multiple time-series channels of vocal parameters. Each matrix contains correlation or covariance coefficients between the channels at multiple time delays. Changes over time in the coupling strengths among the channel signals cause changes in the eigenvalue spectra of the channel-delay matrices. The matrices are computed at multiple time scales corresponding to separate subframe spacings. Features at each time scale consist of the eigenvalue spectra of channel-delay correlation matrices, covariance power (logarithm of the trace), and entropy (logarithm of the determinant) from channel-delay covariance matrices. Under various conditions, we find different degrees of dynamical complexity both within (e.g., formant frequencies, or HNR versus CPP) and across (e.g., formant frequencies vs creak), dependent on the condition, as reflected in eigenvalue distributions. This methodology is illustrated in Figure 6 with the example of the generation of formant track correlation matrices. Further mathematical details of this approach are in Williamson, Bliss, et al. [58],

(a) (b) (c)
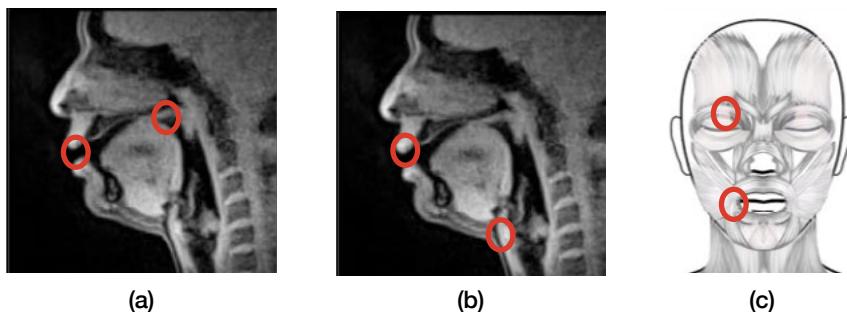
FIGURE 5. Circled in red, the anatomical regions where coordination and timing features are derived are within the articulatory elements that form the vocal tract (a), across vocal tract and vocal source, or fold, components (b), and across muscle groups in facial expression (c). MRI IMAGES COURTESY OF UNIVERSITY OF SOUTHERN CALIFORNIA.

which introduced the method for the analysis of EEG signals for epileptic seizure prediction, and in Quatieri, Williamson, et al. [43], which describes the approach in the context of speech and face processing.

We have also introduced a feature set that characterizes the structure of signal coherence and power at multiple frequency bands within and across speech components over time. We refer to this as coherence structure. The coherence between channels, indicating the amount of cross-channel power in a frequency band relative to the amount of within-channel power, provides a measure of how closely related the signals are within a frequency band. The power and cross-power are computed among three formant frequency channels in three different frequency bands, and a 3×3 coherence matrix is constructed for each band. Analogous to correlation structure, the eigenspectra of the coherence matrices indicate the structure of coherence across the channels.

Our high-level vocal source and prosodic features rely on their low-level counterparts described earlier. In one case, we leverage a finer time scale on a phoneme level (within phoneme boundaries), using feature sets of phoneme-dependent durations [56]. Based on estimated average durations for each phoneme, the summed average durations of certain phonemes are linearly combined to yield fused phoneme duration measures. A subset of phonemes whose summed durations are highly correlated with an assessment (e.g., known depression or cognitive load level) on a training set is selected to create these fused measures, with weights based on the strength of their individual correlations. This phoneme-dependent feature paradigm is shown in Figure 7. A fused phoneme-dependent pitch slope measure is also obtained by using essentially the same procedure as described above. For each passage in a training set, the set of phonemes with the highest correlating summed pitch slopes is selected.

Facial Features

As with vocal features, high-level facial features rely on low-level features. Likewise, analogous to vocal high-level features, we use correlation-based measures that reflect the coordination of facial muscle groups and a rate measure that reflects the duration of each muscle group.

LOW-LEVEL FEATURE EXTRACTION

Characterizing the effects of neurological disorders on facial movements is an active research area. For example, among people suffering from major depressive disorder, measurable differences have been found in facial expressions [22], including acute reductions in involuntary facial expressions in depressed persons [23] and changes in facial expressions that are imperceptible to clinicians [18]. The facial action coding system (FACS) is a systematic method for quantifying localized components of facial expressions called facial action units (FAUs). Each FAU corresponds to distinct sets of muscle movements of the face [24]. The University of California San Diego has developed a computer expression recognition toolbox (CERT) that provides automatic identification of FAU likelihoods or probabilities from individual video frames [27]. Figure 8 lists the FAUs output by CERT for video-based facial expression analysis.

HIGH-LEVEL FEATURE EXTRACTION

The coordination of facial movements during speech can be measured by using correlation structure features obtained from the multivariate FAU time series described in Figure 8. With the same method we used to assess coordination from audio-based features, we can construct high-dimensional channel-delay correlation matrices in which each matrix element represents the correlation coefficient between two FAU time series (channels) at a particular relative time delay [54]. Analogous to our
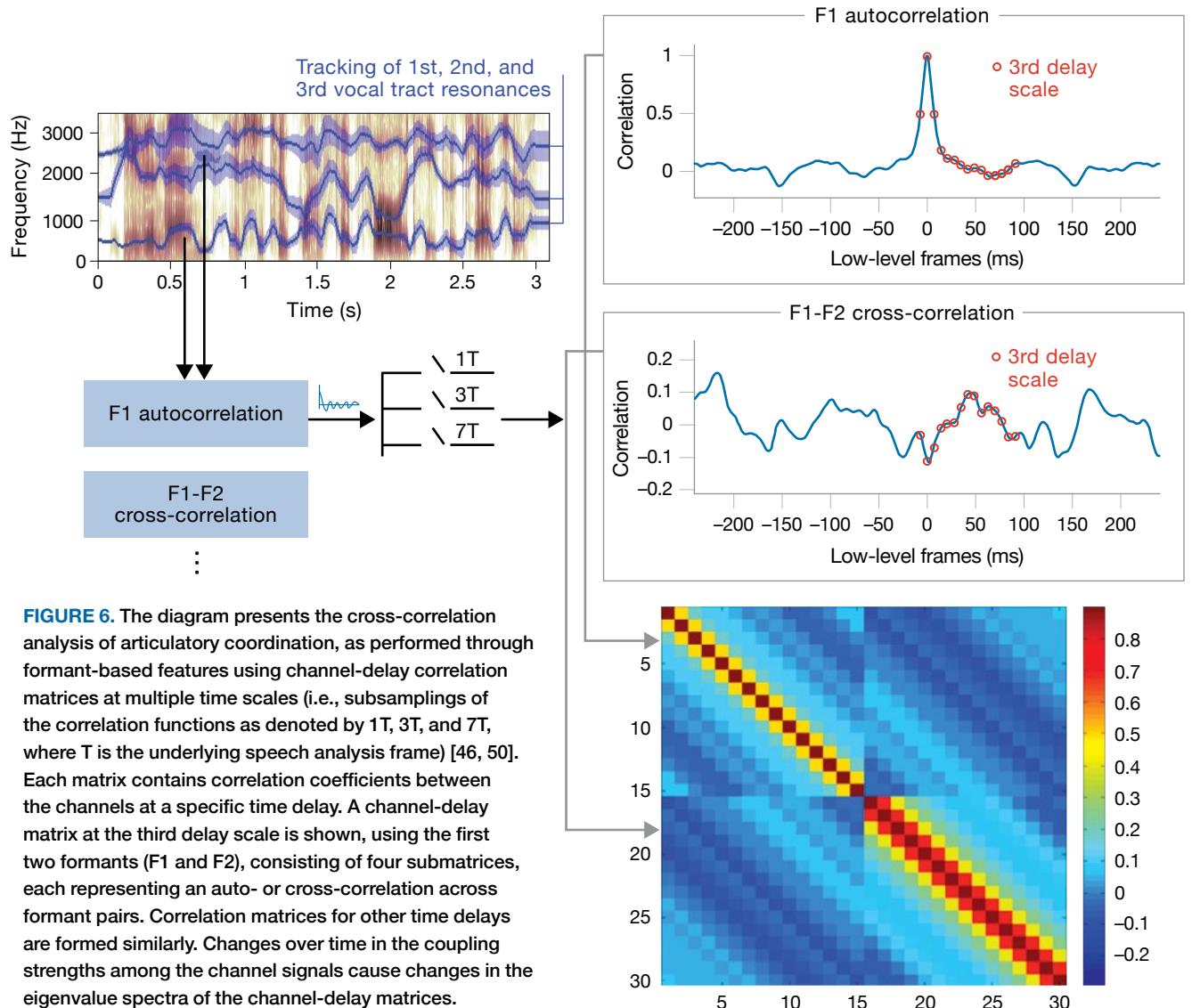
FIGURE 6. The diagram presents the cross-correlation analysis of articulatory coordination, as performed through formant-based features using channel-delay correlation matrices at multiple time scales (i.e., subsamplings of the correlation functions as denoted by 1T, 3T, and 7T, where T is the underlying speech analysis frame) [46, 50]. Each matrix contains correlation coefficients between the channels at a specific time delay. A channel-delay matrix at the third delay scale is shown, using the first two formants (F1 and F2), consisting of four submatrices, each representing an auto- or cross-correlation across formant pairs. Correlation matrices for other time delays are formed similarly. Changes over time in the coupling strengths among the channel signals cause changes in the eigenvalue spectra of the channel-delay matrices.

vocal measures, the eigenspectrum of the correlation matrix then characterizes the total level of independent movement (i.e., complexity) contained in the FAU set.

The facial coordination features represent the total amount of independent movement captured by the FAUs, without regard to the level of movement contained in individual FAUs. Facial activation rate features take a different tack, representing the average level of activation among those FAUs that correlate strongly with the outcome measure of interest, such as depression score [54]. Analogous to our phoneme-duration strategy, an aggregate measure of facial activation rate is then obtained by linearly combining the rates of the highly

correlated FAUs, with negative weights assigned to FAUs with negative correlations.

Electroencephalogram Features

Although not often feasible for a long-term objective of a nonintrusive mobile platform, electroencephalogram (EEG) is used as a reference and sometimes as ground truth. As with vocal and facial characterization, we work with low- and high-level features in EEG analysis.

The EEG signals were measured with a 1000 hertz (Hz) sampling with a 64-element Neuroscan system, followed by high-pass filtering and standard artifact removal. The EEG signals were decomposed into five

frequency bands that have been implicated in cognitive, sensorimotor, and perceptual activities (delta, theta, alpha, beta, gamma), with band ranges of 0–4 Hz, 4–8 Hz, 8–16 Hz, 16–32 Hz, and 32–49 Hz, respectively [25].

Decomposition was performed by bandpass filtering in each of these frequency bands at each channel. In our low-level analysis, we use EEG channel-dependent band power to compute measures of spatial activity patterns, which provide a basis for and complementary information to high-level features.

In high-level analysis, our guiding principles are that successful cognition requires coordinated neural activations in brain networks linking multiple brain regions and that these networks communicate using oscillatory codes operating over a wide range of frequencies. Based on these principles, our high-level feature approach is to use measures of neural coordination indicated by EEG connectivity at each frequency band. We use two connectivity measures: pairwise channel coherence and covariance [43, 67]. Coherence measures cross-channel power relative to within-channel power, whereas covariance measures cross-channel correlation weighted by within-channel power.

Neurological Disorders

The general classification framework and biomarker extraction approach have been applied to numerous psychological and neurological disorders. Researchers in the Human Health and Performance Systems Group at Lincoln Laboratory have applied these principles to the detection of major depressive disorder, Parkinson's disease, mild traumatic brain injury, dementia, and amyotrophic lateral sclerosis (ALS). Here, we focus on the first three of these conditions.

Predicting Major Depressive Disorder Severity

Major depressive disorder (MDD) is the most prevalent mood disorder, with a lifetime risk of experiencing the disorder ranging from 10 to 20 percent for women and 5 to 12 percent for men [68]. As the number of people suffering from MDD steadily increases, so too does the burden of accurate diagnosis. Currently, the diagnosis of MDD requires a comprehensive assessment by a professional with significant clinical experience. However, the inter-clinician variability of these assessments makes the tracking of medication efficacy during clinical trials
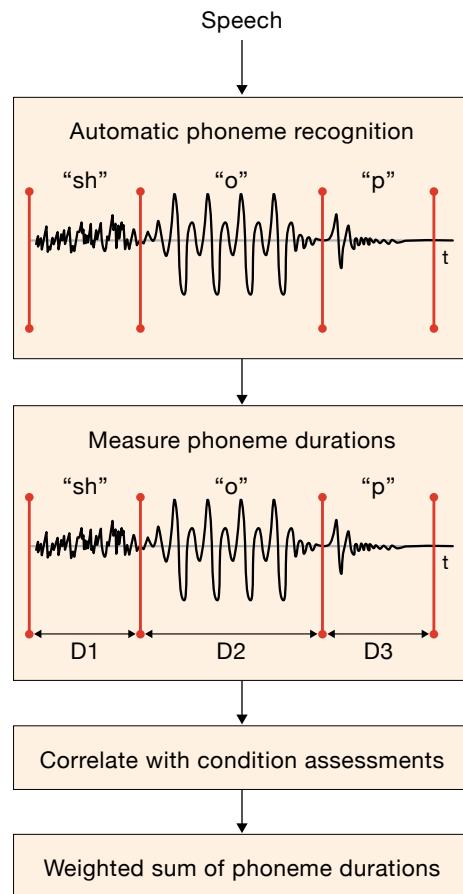
FIGURE 7. Phoneme-dependent duration extraction [56] first requires automatic phoneme recognition, followed by measuring the average duration (D) of each phoneme (there are 42 in the English language). Average duration measures are then correlated with the condition assessment (e.g., severity of depression) across subjects. The final feature is a weighted sum of phoneme durations, where weights are a function of the correlation value for each phoneme.

difficult. The growing global burden of MDD suggests that a convenient and automated method to evaluate depression severity would both simplify and standardize the tasks of diagnosing and monitoring depression, allowing for greater availability and uniformity in assessment. An automated approach may reduce multiple in-office clinical visits, facilitate accurate measurement and identification, and quicken the evaluation of treatment. Toward these objectives, potential depression biomarkers of growing interest are vocal and facial expression features, two categories of easily acquired measures that have been shown to change with a patient's mental condition and emotional state [14–16, 24, 27, 28, 66, 69].

ACTION UNIT NUMBER	FACIAL ACTION UNIT DESCRIPTION
1	Inner brow raiser
2	Outer brow raiser
4	Brow lowerer
5	Upper lid raiser (eye widen)
6	Cheek raiser
7	Lid tightener
9	Nose wrinkler
10	Upper lip raiser
12	Lip corner puller
14	Dimpler
15	Lip corner depressor
17	Chin raiser
18	Lip pucker
20	Lip stretcher
23	Lip tightener
24	Lip presser
25	Lips part
26	Jaw drop
28	Lip suck
45	Blink/eye closure

(a)
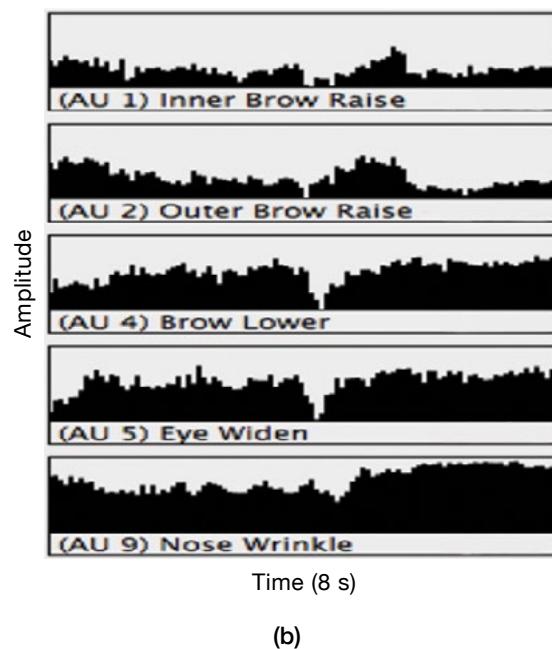
(b)

FIGURE 8. The 20 facial action units from the computer expression recognition toolbox, or CERT, are listed in (a) with their corresponding action unit (AU) number. (We list only the 20 facial action units that are included in CERT; there are many additional facial action units not listed here.) Example time series of five of these facial action units, extracted from video during speaking, are shown in (b). The time series illustrate time variations of each of these facial action units.

VOCAL AND FACIAL CHARACTERIZATION

Depression exhibits changes in all three vocal components described above: speech excitation (source), vocal tract (system), and pattern of stress and intonation (prosody). Depression-related changes in speech reflect the perception of qualities, such as monotony, slur, slowness, hoarseness, and breathiness, in the speech of depressed individuals. Hoarseness and breathiness may be associated with speech source characteristics (at the level of the vocal folds). Monotony may be associated with prosody (e.g., modulation of speech-rate, pitch, and energy),

and slur with speech system characteristics (e.g., vocal tract articulators). Likewise, characterizing the effects of depression on a speaker's facial movements is an active research area. Early work found measurable differences between facial expressions of people suffering from MDD and facial expressions of nondepressed individuals [24]. Electromyography monitors can register facial expressions that are imperceptible during clinical assessment [18] and have found acute reductions in involuntary facial expressions in depressed persons [23].

Although significant effort has focused on studying vocal and facial biomarkers for emotion classification, little or no study has been done investigating changes in coordination, movement, and timing by using speech and facial modalities for depression classification or severity prediction. In individuals suffering from MDD, neurophysiological changes often alter motor control and thus affect mechanisms controlling speech production and facial expression. Clinically, these changes are typically associated with psychomotor retardation, a condition of slowed

neuromotor output manifested in altered coordination and timing across multiple observables of acoustics and facial movements during speech [53]. Alluded to earlier is evidence that a nonspeech region called the limbic system (that controls mood and emotion) is disrupted, and this disruption propagates to the core speech network and in particular neural circuits that control timing and coordination of the articulators (Figure 9). Consistent with a disruption of timing and coordination circuits is the general observation by clinicians that depressed subjects, on the average, tend to talk more slowly and have less clarity in their articulation than nondepressed individuals. Likewise, during speaking, the corresponding facial expression is slower and flatter in effect.

FEATURE SELECTION

Motivated by this neural-based hypothesis and clinical observations, we developed and applied a variety of the high-level articulatory and facial coordination features and vocal and facial timing features outlined earlier [46, 53, 54, 56]. Our high-level coordination-based features are designed to characterize properties of coordination from the low-level features. After investigating multiple combinations of the low-level vocal features as input to correlation analysis, we found the best overall performance was achieved by using the following three combinations: (1) formant–CPP, (2) CPP–HNR, and (3) delta MFCC. Channel-delay correlation and covariance matrices are computed from multiple time-series channels (of given vocal and facial parameters). Each matrix contains correlation or covariance coefficients between the channels at multiple relative time delays. Changes over time in the coupling strengths among the channel signals cause changes in the eigenvalue spectra of the channel-delay matrices. The matrices are computed at four separate time scales, in which successive time delays correspond to frame spacings of 1, 3, 7, and 15. Overall covariance power (logarithm of the trace) and entropy (logarithm of the determinant) are also extracted from the channel-delay covariance matrices at each scale. For vocal-based timing features, we use cumulative phoneme-dependent durations and generalize to phoneme-dependent pitch slopes, obtained by using estimated phoneme boundaries. For facial-based timing features, we use FAU rates obtained from their estimated posterior probabilities.

EXAMPLE DETECTION

We have tested our biomarkers on a variety of depression databases. Here we focus on the 2014 Audio/Video Emotion Challenge (AVEC) that uses a depression corpus that includes audio and video recordings of subjects performing two human-computer interaction tasks in the German language: (1) reading a phonetically balanced passage and (2) replying to free-response questions [54]. Data were collected from 84 German subjects with ages ranging between 18 and 63 years, with a mean of 31.5 years and a standard deviation of 12.3 years. Video of the subjects' faces was captured with a webcam at 30 frames per second and a spatial resolution of 640 × 480 pixels. Audio was recorded with a headset microphone at a sampling rate of 32 kHz or 48 kHz. For each session, the standard self-reported Beck Depression Inventory (BDI) assessment score was available. The recorded sessions were split into two datasets: 100 subjects in the training set for designing a classifier and 50 subjects for the test set.
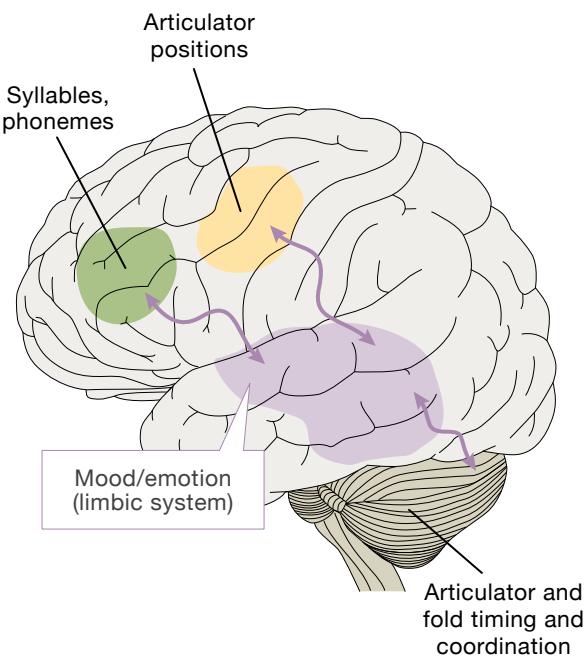
FIGURE 9. A simplified view of the modulation of representative components of the core speech network by the limbic system in depression is illustrated. Our hypothesis is that depression disrupts the limbic system, modifying core production regions required for precise timing and coordination of sound production. This hypothesis is supported by observed average slowing of speaking rate and general decline in clarity.

For our coordination-based features, Figure 10 shows examples of eigenvalue characterization from free-speech data with eigenspectra of the depressed subjects containing less power in the small eigenvalues, indicating a lower level of independent movement. This effect was observed across a spectrum of BDI scores from 83 different subject recordings in which the averages of normalized eigenvalues are plotted for different BDI score ranges (a higher score range indicates greater depression severity): 0–8 (blue), 9–19 (cyan), 20–28 (green), and 29–45 (red). Three different feature types are illustrated: vocal tract control (formants), clarity (vocal source: HNR-CPP), and facial movement (FAUs). A monotonic

decrease in the average normalized eigenvalues with higher BDI score indicates that depression correlates with lower levels of independent vocal and facial movements, i.e., lowered ability for complex motor control.

For our timing features, Table 1 shows cumulative phoneme-dependent durations, obtained from estimated phoneme boundaries; for facial-based timing features, we see FAU rates obtained from their estimated posterior probabilities. Specifically, Table 1 shows the 10 highest correlations with BDI scores of average phoneme durations (a) and FAUs (b) from the AVEC dataset for speech from a read passage (North Wind) and from free speech. For facial-based timing features, we use FAU rates obtained from their estimated posterior probabilities. Notice that the aggregate measures have higher correlations than those of any individual feature type.

Our next step involved mapping the features into univariate scores that can be easily mapped into BDI predictions. To do this, we used a novel extension of a generative Gaussian mixture model (widely used for automatic speaker recognition [65]) referred to as a Gaussian staircase classifier [54], and a discriminative extreme learning machine, a single layer feedforward neural network architecture with randomly assigned hidden nodes [57].

Our overall prediction system is shown in Figure 11. In both training (construct models) and testing (apply

(a)

(b)

(c)
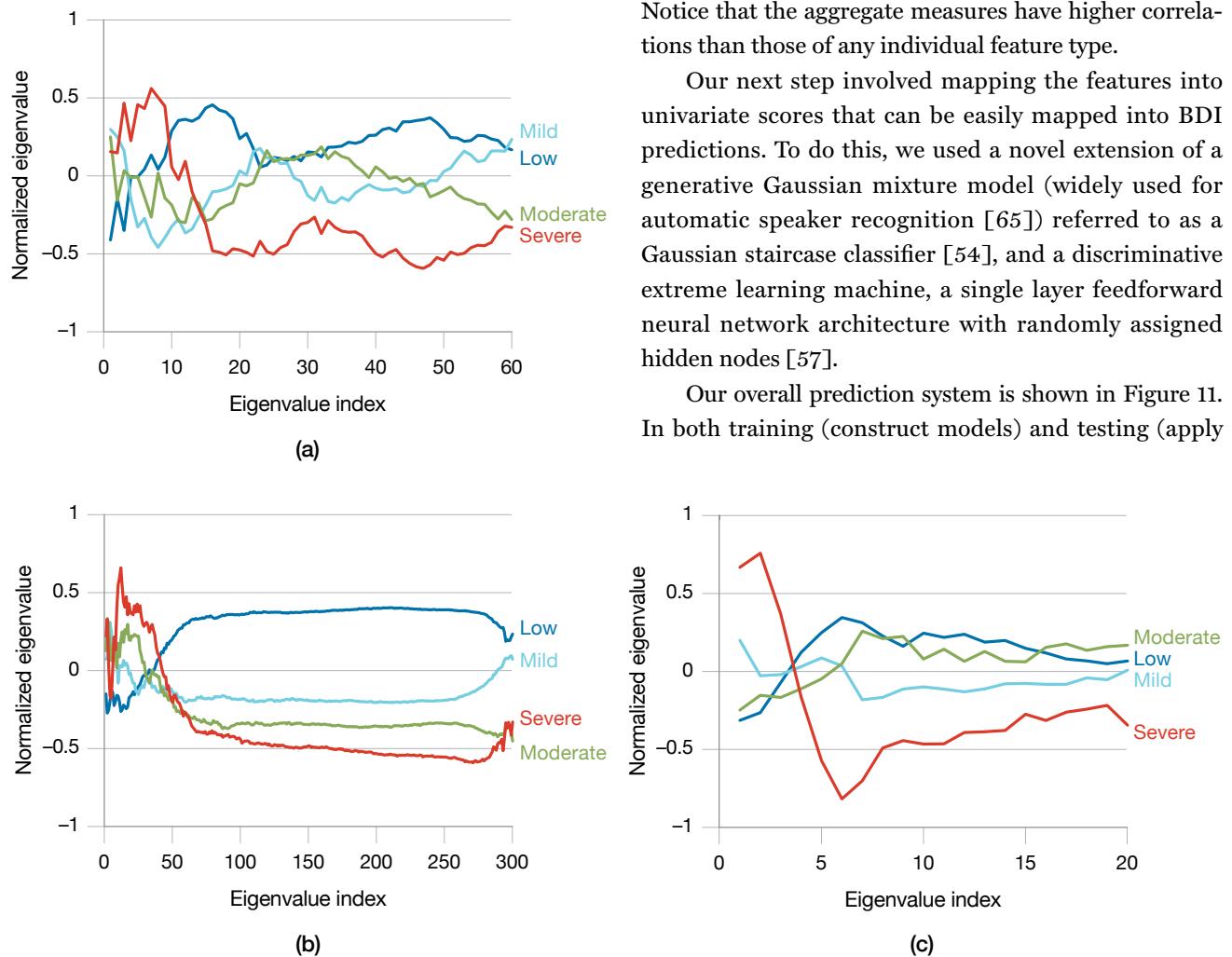
FIGURE 10. Correlation structure features are given as average normalized eigenvalues for four Beck Depression Inventory (BDI) ranges for three different feature types: (a) vocal tract control (formants), (b) facial movements (facial action units), and (c) speech clarity (vocal source: HNR-CPP). The different BDI score ranges are 0–8 (blue), 9–19 (cyan), 20–28 (green), and 29–45 (red). A higher score indicates greater depression severity.

models), after low- and high-level audio- and video-based feature extraction, we typically performed feature dimensionality reduction to account for an often high-dimensionality feature set. From the audio- and video-based features extracted, eight different feature sets were defined. Initial BDI predictions were then obtained from three predictors, which used different combinations of the eight feature sets and two types of classifiers. Each of the three predictions was obtained by using a univariate regression model created from the training set and applied to the classifier output from the test data. The outputs of the three predictors were fused to create a final BDI prediction, using weights based on each predictor's accuracy. Details of the methodology are found in Quatieri, Williamson, et al. [43] and Williamson, Quatieri, et al. [54].

The prediction system shown in Figure 11 was used for our best submission in the AVEC 2014

Challenge, with BDI score root-mean-squared error (RMSE) = 8.12 across the test data. A similar system was used in the AVEC 2013 Challenge. In both challenges, Lincoln Laboratory took first place. Note that an alternative objective is detection of depression state. We can achieve this through mapping BDI scores into two ranges—severely depressed (25–50 BDI score range) and moderately depressed (0–24)—and then mapping predicted scores to one of the two classes to form a detector. Figure 12 shows the receiver operating characteristics (ROCs) (false-alarm versus true detection rates) for this mapping in both the AVEC 2013 and 2014 Challenges.

Predicting Parkinson's Disease Severity

Parkinson's disease is a neurological disorder with associated progressive decline in motor precision and

Table 1. Correlation Coefficient (*r*) of Average Phoneme Durations (a) and Facial Action Unit (FAU) (posterior probabilities) Timing Features (b) with Beck Depression Inventory Scores

NORTH WIND		FREE SPEECH			NORTH WIND		FREE SPEECH	
Phoneme	*r*	**Phoneme**	*r*		**FAU**	*r*	**FAU**	*r*
n	−0.38	ih	0.34		Brow lower	0.24	Brow lower	0.22
ah	0.35	ey	−0.27		Dimpler	−0.28	Dimpler	−0.18
d	−0.23	r	0.25		Eye widen	−0.26	Eye widen	−0.27
ih	0.22	sil	0.24		Lip stretch	−0.24	Lip stretch	−0.15
b	−0.20	oy	−0.20		Lip tightener	−0.30	Lip tightener	−0.22
g	0.18	ah	0.19		Nose wrinkle	0.21	Nose wrinkle	0.19
ae	0.17	y	0.16		Chin raise	−0.23	Cheek raise	0.14
dh	−0.15	m	0.16		Lip corner pull	−0.30	Lids tight	0.16
eh	0.15	g	−0.14					
iy	−0.15	er	0.14		Lip pucker	0.27	Lip corner depressor	0.17
Fused	0.63	Fused	0.51		Jaw drop	0.37	Outer brow raise	−0.16
					Fused	0.58	Fused	0.46

(Common FAUs for both passages: Brow lower, Dimpler, Eye widen, Lip stretch, Lip tightener, Nose wrinkle)

(a)

(b)

For each session in the training set, the subject's self-reported Beck Depression Inventory assessment score was available.

sensorimotor integration, stemming presumably from the basal ganglia. In this disorder, a steady loss of cells in the midbrain leads to speech impairment in nearly 90 percent of subjects [19]. Early, accurate detection of Parkinson's disease may aid in possible intervention and rehabilitation. Thus, as with MDD, simple noninvasive biomarkers are desired for determining disease severity. Toward this end, we have applied a methodology similar to that developed for predicting MDD, introducing a novel set of acoustic speech biomarkers and fusing them with conventional features to provide clinical assessment of Parkinson's disease.

VOCAL CHARACTERIZATION

Our acoustic biomarkers reflect changes in speech production that are due to disturbances in underlying neurophysiology that affect the source, system, and prosodic components of vocal expression. In particular, impairment to the basal ganglia may in turn modulate (and cause impairment in) the core speech network, analogous to the interaction shown in Figure 9 with depression. Speech and voice characteristics of Parkinson's disease include imprecise and incoordinated articulation, monotonous and reduced pitch and loudness, variable speech rate and rushes of breath and pause segments, breathy and harsh voice quality, and changes in intonation and rhythm [19, 30, 31, 57, 70, 71]. Such changes occur at phonetic and larger time scales, including multiscale perturbations in formant frequency and pitch trajectories, in phoneme durations and their frequency of occurrence, and in temporal waveform structure. We have also introduced articulatory features based on a neural computational model of speech production that is introduced in more detail in the article "Fundamental Brain Research" in this issue of the *Lincoln Laboratory Journal*.
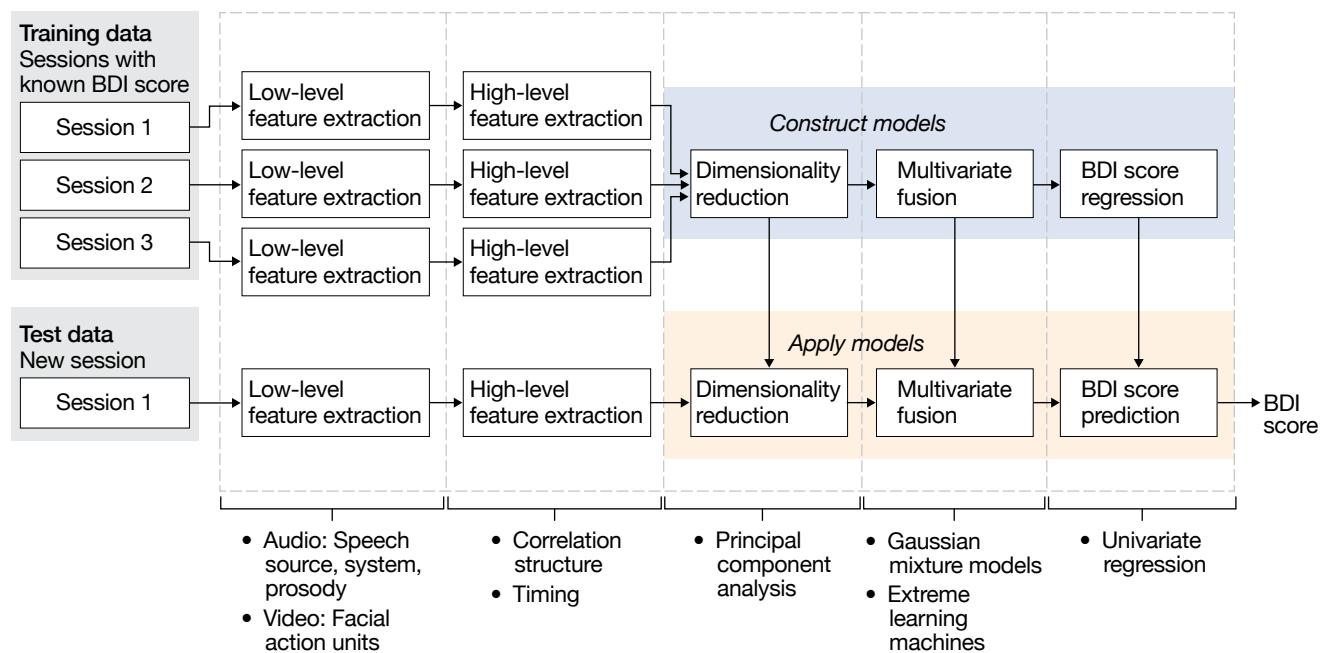
FIGURE 11. This figure shows a block diagram of the developed system for predicting the Beck Depression Inventory (BDI) assessment score. In training to obtain classifier models and in testing, the processing pipeline involves extraction of high-level features derived from low-level features, dimensionality reduction, fusion, and BDI score regression based on different classifiers. Low-level feature extraction characterizes properties of the speech or video signal within short duration (10 ms or 30 ms) time windows. High-level feature extraction computes summary statistics from these low-level features over an entire passage of recorded speech. These high-level features are composed of the eigenspectra of correlation matrices, which are created using time-delay embedding of the low-level feature time series at multiple time-delay scales. Dimensionality reduction of these eigenvalue-based feature vectors is done by using principal components analysis. Finally, Gaussian mixture models and extreme learning machines are used to produce prediction scores that are mapped into BDI predictions by using univariate regression.
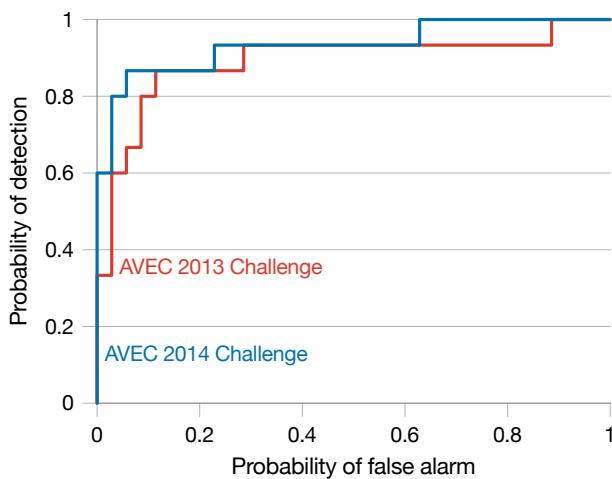
FIGURE 12. ROC detection curves (false alarm versus true detection rates) are shown for the AVEC 2013 and 2014 Challenges. In 2013, only speech was used, while in 2014, both speech and facial action units were leveraged but with considerably less training data provided. In both cases, true detection versus false-alarm rate trade-offs were favorable, i.e., for about 20 percent false alarms, we obtain about 85 percent detection accuracy. An automated system that can classify major depressive disorder at this level of accuracy would be extremely useful as a clinical screening tool.

FEATURE SELECTION

Our feature development reflects the three basic aspects of speech production: phonation (source), articulation (vocal tract), and prosody (intonation and timing). Our methodology is largely similar to those used in predicting depression severity: we design high-level features to characterize properties of timing and coordination from the low-level features. For example, our primary features exploit changes in phoneme-dependent durations, pitch slope, and formant frequency slope. The effectiveness of these features is consistent with previous findings that certain speech segments are more prone to variation than others in Parkinson's disease [72, 73].

We also use a high-level correlation structure of formant trajectories but expand this concept to a correlation structure of the position of speech articulators derived from a neurocomputational model of speech production, the Directions into Velocities of Articulators (DIVA) model [48]. (See the article "Fundamental Brain Research" in this issue for a more complete description of this approach, which uses a vocal source model in contrast to a vocal tract system model, and its application to depression.) The DIVA model takes as inputs the

first three formants and the fundamental frequency of a speech utterance. Through an iterative learning process, the model then computes synaptic weights that correspond to modules of the speech production process, including aspects of the articulatory feedforward mechanism and auditory and somatosensory feedback errors (Figure 13). We hypothesize that Parkinsonian speech results from impairments in certain components of the speech production process, and therefore, when the model is trained on Parkinsonian speech via the iterative learning process, the internal variables reflect the severity of the disorder [57]. In this work, we have focused specifically on the correlation structure features derived from the DIVA model's 13 time-varying articulatory position states, with the same delay and scale parameters used for the correlation structure of the formant trajectories. More details and the complete feature selection process are described in Williamson et al. [57].

EXAMPLE DETECTION

One Parkinson's disease database we are working with is from the Interspeech 2015 Computational Paralinguistic Challenge described in Orozco-Arroyave et al. [19]. Assessments of Parkinson's severity are based on the Unified Parkinson's Disease Rating Scale (UPDRS) with a score range of 6–92 [35]. A higher score indicates higher severity. The language is Spanish and the dataset is divided into 42 tasks per speaker, yielding 1,470 recordings in the training set (35 speakers) and 630 recordings in the development set (15 speakers), both with UPDRS scores provided. The dataset also contains 462 recordings (11 speakers) in the test set, without UPDRS scores provided. The duration of recordings ranges from 0.24 seconds to 154 seconds.

As an example of the discriminatory capability of our articulatory coordination–based features derived from the DIVA neurocomputational model, Figure 14a shows eigenvalue characterization of the correlation structure features derived from the model's 13 time-varying articulatory position states. The average eigenvalues for three different ranges of Parkinson's severity are computed (in standard units) for the sentence "Luisa rey compra el colchón duro que tanto le gusta" across speakers from the training set. These averages reveal distinct differences related to Parkinson's disease severity. These

FIGURE 13. Biomarkers are derived from a neurocomputational speech production model, which includes articulatory feedforward mechanisms and auditory and somatosensory feedback errors. The model's optimal parameter fit, given a measured speech signal, provides input to high-level feature extraction by using correlation structure analysis.
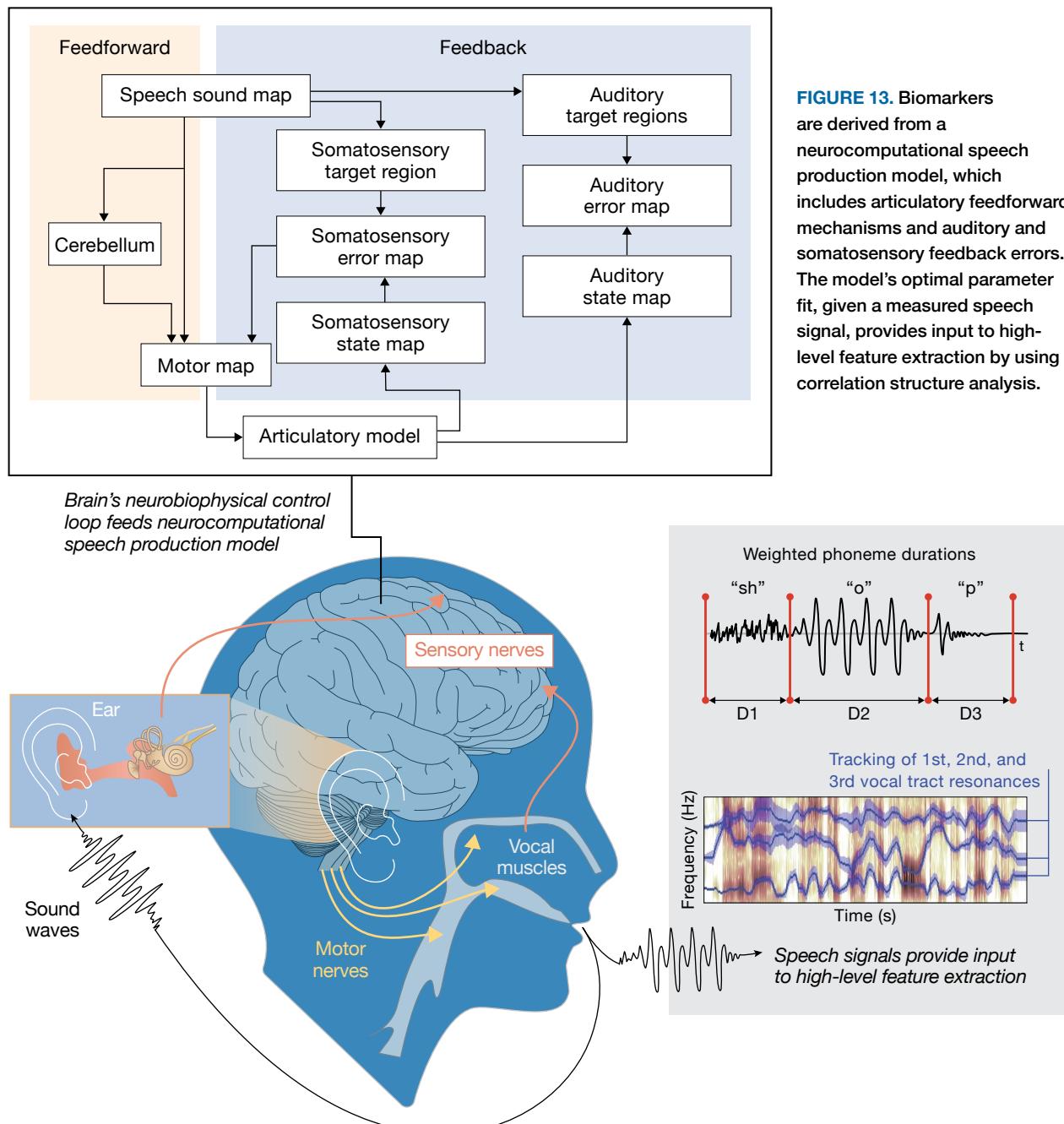
differences were observed across a spectrum of scores for correlation structure features from different tasks. The decrease in the higher-indexed average normalized eigenvalues with higher Parkinson's disease score indicates that Parkinson's disease correlates with lower levels of independent vocal movements, i.e., lowered ability for complex motor control, as we observed in major depressive disorder.

In an approach similar to our detection prediction methodology, we fused a subset of our set of speech features into the Parkinson's severity prediction system by using multiple Gaussian staircase statistical regression models. Using a linear combination of the predictions from the multiple statistical models, we obtained Spearman correlations (a nonparametric measure of dependence between two variables) between predicted

scores and clinical assessments (the metric of choice for the Interspeech 2015 database) of $r = 0.63$ on the training set (fourfold cross validation), $r = 0.70$ on a held-out development set, and $r = 0.96$ on a small held-out test set [57]. As illustrated in Table 2, our neurological model–based features provide a sizable gain over our more standard timing and coordination features.

Using a system analogous to our depression prediction system, we can split the UPDRS scores in the middle and design a binary detection problem. Figure 14b shows the resulting true detection versus false-alarm rate for this binary detection scenario, indicating that the algorithm can effectively discriminate between minimal to mild and moderate to severe Parkinson's disease levels. Features used in detection are derived from a combination of primarily phonetic timing and formant coordination representations and articulatory coordination derived from the DIVA neurocomputational model [57].

Detecting Cognitive Decline from Neurotrauma

Mild traumatic brain injury (mTBI) affects an estimated 1.7 million civilians each year, who incur the condition primarily because of sport injuries, falls, and car accidents. Servicepeople are at significant risk of mTBI because of the hazards associated with military training and operations. A wide variety of sensorimotor (e.g., speech, dexterity, vestibule) and cognitive (e.g., attention, memory) problems may result from mTBI. Changes in neurocognitive status following concussion injury may adversely impact an individual's work and daily life, and, in the military context, may affect individual and unit readiness. The ability to determine cognitive performance changes associated with mTBI in an objective, noninvasive way would facilitate the monitoring of injury during all stages of care and recovery, and would serve as a valuable decision aid to leaders for determining when a soldier can safely return to duty.

Since speech and cognitive processes are highly coupled in the brain (sharing similar pathways and regions, perhaps each modulating the other), we would expect to be able to detect cognitive changes through changes in vocal and facial expression during speaking. We applied the methodology introduced earlier to a civilian case (high school athletes) and a military case (U.S. veterans), showing that our features based on timing and coordination of the complex motor activity

(a)

(b)
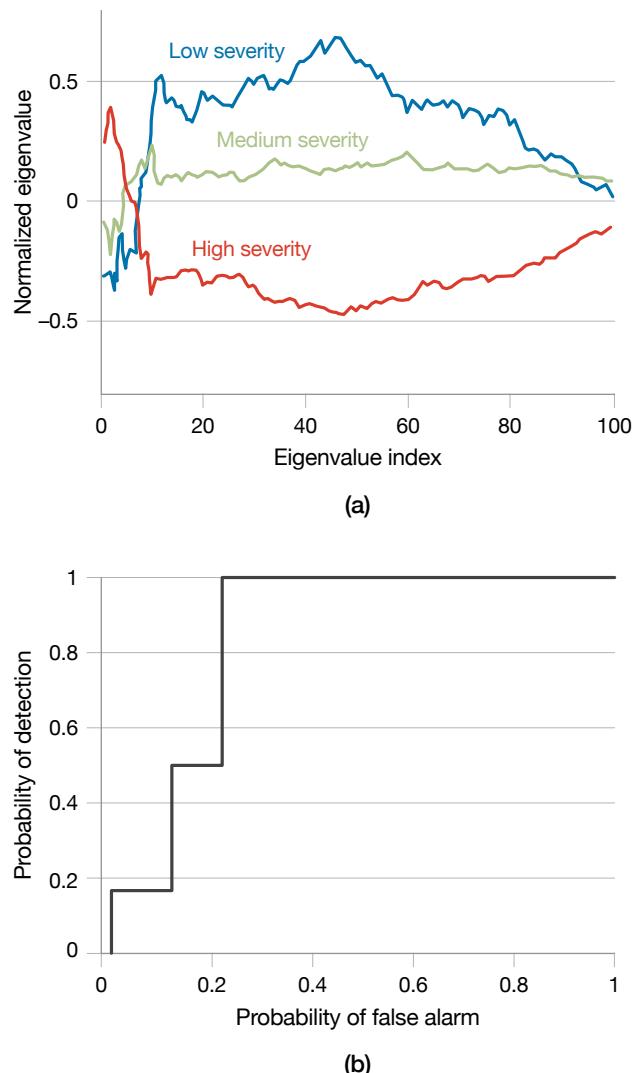
FIGURE 14. The plot in (a) shows average eigenvalue features on the development set for three ranges in the Unified Parkinson's Disease Rating Scale: 0–20 (blue), 21–40 (green), and 41–100 (red). The plot in (b) shows the true detection versus false-alarm rate for this detection scenario on the test set, indicating that the algorithm can effectively discriminate between minimal to mild and moderate to severe Parkinson's disease levels. Features used in detection are derived from a combination of primarily phonetic timing and formant coordination representations and articulatory coordination derived from the DIVA neurocomputational model [57].

of vocal and facial expression while speaking provide sensitive indication of cognitive impairments resulting from neurotraumatic injury. The datasets we explored represent a variety of mTBI types, some involving concussion and some requiring immediate evaluation and others requiring longer-term evaluation. Even when

THOMAS F. QUATIERI, JAMES R. WILLIAMSON, ADAM C. LAMMERT, KRISTIN J. HEATON, AND JEFFREY S. PALMER

Table 2. Sensitivity Improvement Plus Ability to Specify Affected Neural Pathways by Using Lincoln Laboratory Neurological Model–Based Features

	STANDARD LINCOLN LABORATORY PHYSIOLOGICAL VOCAL TRACT FEATURES	NEUROLOGICAL MOTOR COORDINATION FEATURES	STANDARD PLUS NEUROLOGICAL MOTOR FEATURES
Correlation metric	0.88	0.39	0.96

concussion does not occur, repeated exposure to head impacts without concussion can cause neurocognitive and neurophysiological impairments prior to concussion (referred to as preclinical mTBI) [37].

VOCAL AND FACIAL CHARACTERIZATION AND FEATURE SELECTION

With brain trauma, a wide range of lesion types, severities, and locations relate to different types and severities of speech and cognitive impairments. Changes in brain structure or connectivity may result in changes in source, prosodic, or articulatory aspects of voice. These changes include excessive delays in initializing a vocalization; excessive (exaggerated) or reduced (flat) emotional content; impaired function of muscles affecting the lips, tongue, vocal folds, and/or diaphragm; and difficulty in making and coordinating the precise movements of these articulators [74]. In our work, we have examined the articulatory components of speech reflected in formant tracks and the ways that changes in track dynamics and coordination map to cognitive decline. We also have focused on the change in average phoneme duration as one component of prosodic change caused by brain injury from trauma. As with formant changes, change in average phoneme duration may be associated with decline in cognitive processes from the trauma. Figure 15 shows schematically a simplified view of the link of these particular brain modules with cognitive processing that is widely distributed throughout the brain.

Our methodology is again largely similar to that used in predicting depression severity: we design high-level features to characterize properties of timing and coordination from the low-level features. Our primary features exploit changes in phoneme-dependent durations and the high-level correlation structure of formant trajectories. To characterize formant dynamics as a second
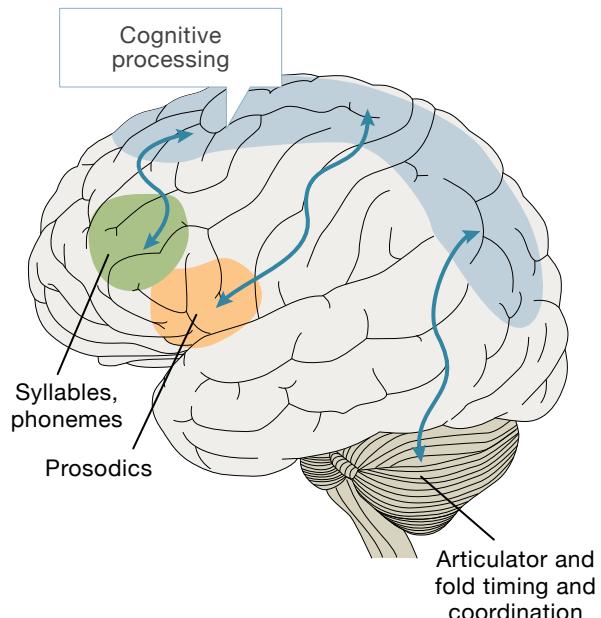
FIGURE 15. A simplified view of the representative brain modules of the core speech production network and their links to higher-level cognitive processing is illustrated. This is a highly schematized illustration given that cognitive processing is broadly distributed throughout the brain.

baseline feature, we extract nine formant functions over 20-millisecond segments at 10-millisecond frame intervals. These functions are the three raw formant tracks and their high-pass and low-pass components. The common 3 dB cutoff frequency of the high- and low-pass filters is 55 Hz. From each of the above functions, we compute three dynamics functions: the raw function value, the velocity, and the acceleration.

DATASETS FOR CASE STUDIES

We hypothesize that preclinical or concussion-related damage results in changes in average vocal tract dynamics measured by formant frequencies,

their velocities, and acceleration; changes in articulatory coordination measured by our formant-frequency cross-correlation characterization; and phoneme-dependent average durations. These features allow machine learning algorithms to detect cognitive changes identified by a battery of cognitive tests. We developed distinct datasets for the two case studies.

1. Athlete study [55, 75]. The dataset for this study involved a population of athletes regularly receiving impacts to the head and showing signs of preclinical mTBI, a state indicated by impaired cognitive performance occurring prior to concussion. Data for this study were collected monthly in collaboration with Dr. Thomas Talavage at Purdue University under a protocol approved by an institutional review board. The study included pre-season, in-season, and post-season data from 32 high school athletes, of whom 25 were male football players and seven were female soccer players. The athletes' ages ranged from 15 to 18, with all data collected independently of any clinical diagnoses of concussions. For each athlete, the data collection included scores from the online Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) assessment version 2.1, which comprises a series of cognitive tests typically used in the sports community [76], along with speech recordings, eye tracking, auditory perception, and optic nerve sheath-diameter measurements [75], all of which are part of a multimodal collection suite developed at Lincoln Laboratory. Although we had obtained promising preliminary results with ocular measures [75], we focused on the speech modality in the athlete study [55]. Speech features were extracted from audio recordings of the Grandfather Passage, which provides a standardized and phonetically balanced sample of speech.

The ImPACT was used as a means of assessing cognitive performance [76]. This test is made up of six subtests that measure verbal memory, visual memory, visual motor speed, reaction time, impulse control, and a total symptom composite. For each test, a threshold is set for a change in cognitive performance. The threshold for each test is defined as a decline from baseline that exceeds one standard deviation, where the standard deviation is computed over the change from baseline across all subjects'

test scores. A support vector machine–based classifier, with cross validation, then uses our formant features to predict a decline in cognitive performance. Figure 16a shows average z-normalized eigenspectra from formant channel-delay matrices associated with cognitive decline (red) and normal function (blue). We compared the effectiveness of vocal tract dynamics features versus articulatory coordination features. This evaluation was done via ROC curves along with a variation of their area under the curve (AUC), where a score of 1 represents a perfect test. For the reaction-time component of ImPACT, the articulatory dynamics features achieved AUC values between 0.72 and 0.98, whereas the articulatory coordination features achieved AUC values between 0.94 and 0.97. Figure 16b illustrates this comparison, indicating the importance of coordination of articulatory components over their absolute counterparts. Nevertheless, for some components of the ImPACT test, e.g., visual motor and verbal memory, the two formant-based features performed about the same.

In this study, we also investigated features reflecting the change from baseline phoneme duration. The features were combined on the basis of their correlation with each of the cognitive modalities and then incorporated into Gaussian classifiers to predict cognitive decline. Classification performance was then analyzed using ROC curves through detection versus false alarm. Using vocal phonetic timing features for the four components of ImPACT that were studied, we computed ROC curves that demonstrated high-fidelity prediction of cognitive change. The highest AUCs achieved were 0.89, 0.80, 0.94, and 0.90 for verbal memory, visual memory, visual motor speed, and reaction time scores, respectively.

2. U.S. veterans study [77]. In this second study conducted in collaboration with Dr. Kristin Heaton at the U.S. Army Institute of Environmental Medicine (USARIEM) and Dr. Alex Lin at Brigham and Women's Hospital, data were collected from 20 subjects. Of the 20 subjects, five had a documented history of mild traumatic brain injury and 15 were control subjects. Subjects were enrolled in a larger study that used magnetic resonance spectroscopy to characterize neurochemical biomarkers for mTBI. Study participants completed a battery of cognitive

tasks selected for sensitivity to changes in cognitive performance associated with mTBI. High-quality audio and video recordings were obtained while participants completed a standardized protocol consisting of read passages, spontaneous speech, and repetition of sounds. Timing and coordination features were extracted from the speech and face data, and used along with select cognitive performance outcomes to construct statistical models for estimating cognitive performance. Speech features were based on articulatory coordination derived from the acoustic signal (vocal-tract resonances), while facial features were derived from the coordination of muscle groups (facial action units).

Gaussian staircase regression models were trained on the 15 control subjects to estimate the processing speed index (PSI) from the Wechsler Adult Intelligence Scale (WAIS-IV), which is sensitive to the cognitive consequences of mTBI. Models were then used to estimate PSI scores from the five mTBI cases not included in the training set. The strongest results were seen in predicting PSI from speech and facial features during the read passages. Pearson's correlation coefficient between the estimates and the recorded PSI scores revealed $r = 0.98$ ($p = 0.003$, $n = 5$) for the speech features and $r = 0.92$ ($p = 0.025$, $n = 5$) for the facial features. Table 3 shows these results in comparison with results for free speech.

These preliminary analyses demonstrate the promise of cognitive assessment technologies based on motor timing and coordination underlying vocal and facial expression during speaking in the context of mTBI.

In the veterans study, statistical models trained on control subjects were transferable to mTBI subjects, while the athletes study addressed accumulative injury over time. Results also show the utility of different modalities for cognitive evaluation following mTBI, with both speech and facial features affording high estimation accuracy in predicting different metrics of cognitive decline. Further validation of these technologies is required on larger datasets and in operational settings. Once validated and fully developed, these assessment technologies may provide a capability for objective, noninvasive real-time assessment of individuals in daily life and in military training and operational settings.
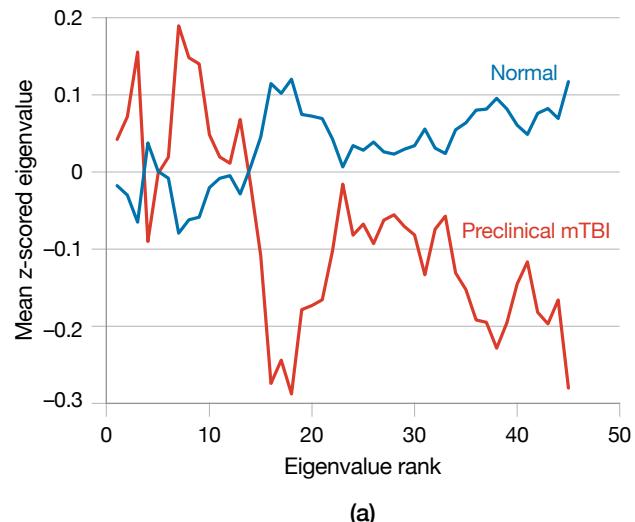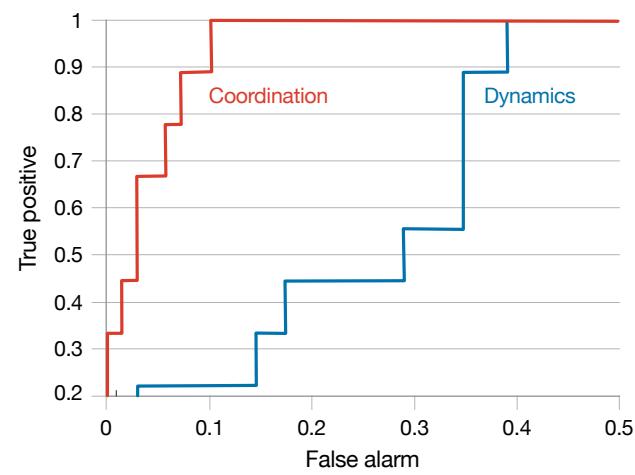
FIGURE 16. The results from the high school athlete study are shown for the reaction-time component of the ImPACT test [55]. Average z-normalized eigenspectra from formant coordination channel-delay matrices associated with cognitive decline are in red and normal function is in blue, showing a large separation especially in certain mid- and high-rank eigenvalues (a). A cognitive decline detection versus false-alarm (ROC) comparison shown in (b) gives the effectiveness of vocal tract dynamics features (blue) versus articulatory coordination features (red).

Cognitive Load

Cognitive load is often defined as the demand placed on cognitive and mental resources required by a particular task [6, 7]. Some tasks place greater burden on such resources and attributes than others. In particular, highly complex tasks, or those involving long periods of sustained activity or attention, and more monotonous

tasks requiring sustained vigilance can tax available cognitive resources and lead to fatigue. An individual's ability to adapt to changing workloads and manage fatigue caused by cognitive exertion can be influenced by stress imposed both externally (e.g., environmental extremes, physical exertion, or social interactions) and internally (e.g., psychological/emotional state, nutritional status, or changes in sleep patterns).

Applications for monitoring cognitive load, including assessing cognitive load, have been developed for both clinical and nonclinical settings. In clinical applications, the objective is often to measure the specific causes of load, while in nonclinical settings, an objective is to quickly assess cognitive ability and readiness under loaded conditions, regardless of their etiology.

Because speech and cognitive processes are highly coupled in the brain (sharing common processes and pathways, and perhaps modulating one another), we might expect to be able to determine cognitive load levels through changes in vocal and facial expression during speaking. We might think of speaking as itself a task that may occur simultaneously with some other specific targeted cognitive task of different load level. Indeed, there is evidence that in such scenarios, dual-task interference occurs under finite cognitive capacity across the dual tasks, with the relevant underlying neural mechanism located in the brain's lateral prefrontal cortex [78]. The linkage may also be due to the coordination of neuromotor activity in the brain and reliance on discrete cognitive functions in the production of speech.

Cognitive Load Detection
Our research has shown that when vocal and facial modalities are combined, they perform nearly as well as a gold-standard EEG analysis in cognitive load detection, thus providing a potential nondisruptive means to track cognitive status.

VOCAL AND FACIAL CHARACTERIZATION
On designing a multimodal database protocol that reflects typical cognitive load conditions, we employed the hypothesis that speech and the corresponding facial movements that occur while speaking are complex motor activities requiring precise neural timing and coordination. We also hypothesized that manipulating cognitive load level systematically alters this complex motor activity

Table 3. Results of the U.S. Veterans Study: Predicting Processing Speed Index [77]

FEATURE SETS	READ SPEECH r (p)	FREE SPEECH r (p)
Formants	0.98 (0.003)	0.45 (0.447)
Facial Action Units	0.92 (0.025)	0.55 (0.332)

Predictions are for five mTBI subjects and from a model trained on 15 control subjects.

in a measurable way [43, 47]. This neural activation is reflected in EEG measurements, which are sometimes considered a gold standard in viewing the effect of working memory demand [11, 12, 79]. To explore those hypotheses, we designed a dual-task protocol involving auditory memory and speaking.

DUAL-TASK PROTOCOL
In one scenario, we introduced cognitive load of different levels through an auditory working memory task shown in Figure 17. Subjects gave informed consent to our working memory–based protocol approved by the MIT Committee on the Use of Humans as Experimental Subjects. Audio data were collected with a DPA acoustic lapel microphone (with a Roland Octa-Capture audio interface), facial video with a Canon high-definition video camera, and EEG signals with a 64-element Neuroscan device.

The working memory task was split into a training and a testing phase. During training, the maximum number of digits that a subject could accurately recall was estimated by using an adaptive tracking algorithm [43, 47]. This number, nc, was used to determine the three difficulty levels in the test phase, which were typically set as: dn = {ceiling(nc), ceiling(nc)-1, ceiling(nc)-2}, where ceiling corresponds to a minimum high load of 4. Despite some minor protocol changes among early subjects, this common load assessment test was used for 10 of the 11 subjects analyzed. We later defined a binary detection problem of discriminating high load (maximum number) from low load (maximum number minus two). The range of the number of digits
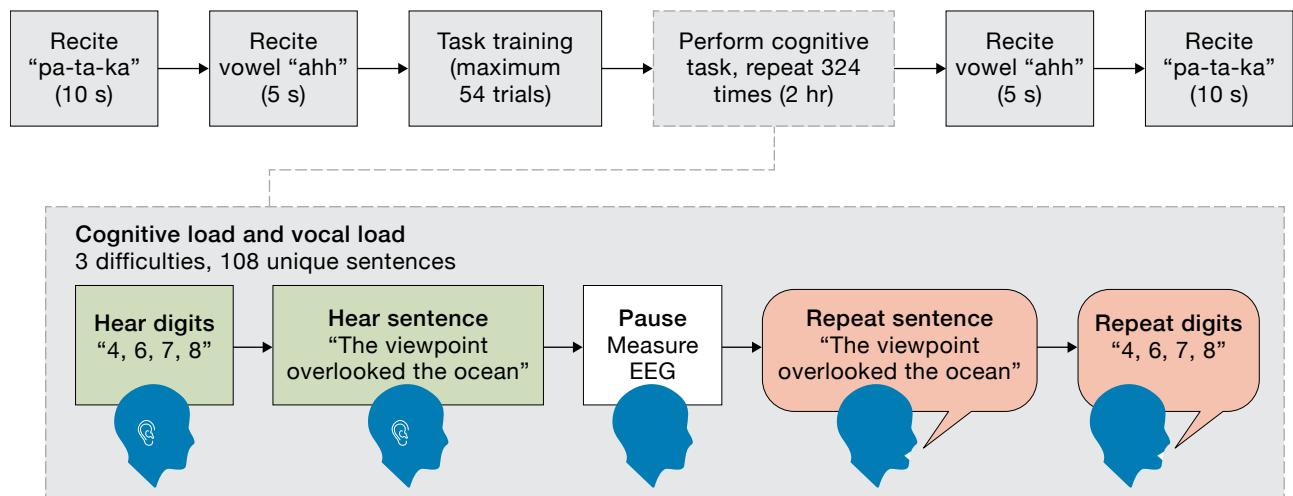
FIGURE 17. Illustrated here is a summary of the working memory protocol. The experiment consists of 324 trials. In each trial, a subject listens to a sequence of digits followed by a sentence (green boxes). Following a pause, the subject repeats the sentence and then the digits (red boxes). In different trials, the working memory load level is modified by varying the number of digits. Electroencephalogram (EEG) is measured during the pause to avoid motion and muscle artifacts. Speech (from audio and facial video) is measured when the sentence is repeated. The working memory load level is estimated based on the EEG and speech signals. Prior to and after the two-hour cognitive task, we introduced reciting the diadochokinetic sequence "pa-ta-ka" and the vowel "ahh," allowing for analysis of speech samples pre- and post-fatigue that are standardly used in articulatory and vocal source analysis.

recalled across all subjects was 2–5 for low load and 4–7 for high load.

Finally, with our protocol, we also measured skin conductance, temperature, and pulse oxygenation level. Though not our focus, the study of these measurements may pave the way to investigating other multimodal biomarkers.

FEATURE SELECTION

Our features from three modalities (voice, face, EEG) are based on the principles of timing and coordination within components of each modality. In each case, we extracted first baseline low-level features followed by our high-level features as functions of the low-level features. For voice, feature vectors were extracted only from the single spoken sentence component of each trial in the test phase of the auditory memory task. Low-level vocal features comprise measures of phoneme and pseudosyllable durations, pitch dynamics, spectral (formant) dynamics, and vocal-fold irregularity (creak). We constructed high-level features that capture timing and interrelationships across the low-level features. The feature sets were derived under the hypothesis that differences in cognitive load produce detectable changes

in speech production timing and coordination within and across articulatory and vocal-fold components. For facial expression, analyzed during the same time interval as audio, the extracted low-level features were facial action units [22, 24, 27], followed by correlation-based measures as high-level features. For the EEG, during the pause interval to avoid motion and muscle artifacts, we performed preprocessing to extract low-level EEG signals free of many typical artifacts, followed by correlation and frequency-dependent coherence and power measures. To avoid motion and muscle artifacts, the EEG measurements were made during time intervals when the subject was not speaking.

As was done in the research on neurological conditions, we averaged across different targeted class conditions. Our high-level features showed strong load discriminability with averages across all subjects of normalized (z-scored) eigenvalues from formant, creak, and delta-MFCC signals for low load and high load. In all three cases, there was greater power in the medium-level eigenvalues during higher cognitive load. This finding indicates greater dynamical complexity in formant frequencies, creak, and spectral content during higher cognitive load and thus higher levels of independent

vocal movements than in the low-load case. Although this finding corresponds to more complex motor control under high load, we hypothesize that the eigenvalue distributions indicate greater independence associated with more random or erratic movements under the high load state—in contrast to our earlier observations for MDD, Parkinson's disease, and mTBI, in which we found less independence of vocal movements. In our cognitive load experiments, we found a similar level of discriminability based on class-conditioned averages of the EEG high-level features, which are coherence eigenvalues of spatial log-power values in the beta frequency band. Also, high load is associated with lower levels of EEG power.

EXAMPLE DETECTION

Although our protocol involves feature processing of single spoken sentences, the ability to detect load after fusing evidence across multiple sentences can be assessed by combining the Gaussian classifier scores from different trials, provided that the trials involve the same load condition. This combining was done by randomly selecting, from the same subject, a number of trials of either high load or low load and summing their Gaussian classifier scores. For each subject, load condition, and combination number, 200 randomly chosen sets of trials were used to determine the fused scores across multiple sentences.

Figure 18 summarizes the ROC results (detection versus false alarm) for the modalities in various combinations. We observe that the EEG-based detector rapidly converges to a limit that, after six minutes, slightly outperforms the combined audio and video modalities that converge to a limit more slowly. Each modality alone (audio or video) converges to a limit that underperforms the combination of audio and video, with audio outperforming video [43] (not shown). Observe that combining all three modalities provides only a small gain over the EEG.

Predicting Cognitive Fatigue

Individuals often report the subjective impression of weakness or slowness in performing cognitive tasks, particularly following periods of heavy cognitive workload. This subjective feeling has been called cognitive fatigue and is closely associated with the objective notion of cognitive capacity (alluded to earlier in the context of dual tasks), which is the ability of an individual

to perform cognitive work. When cognitive workload exceeds an individual's overall capacity, at a given point in time, performance in managing the workload can be expected to degrade. For individuals engaged in tasks such as driving a car or piloting an airplane, reductions in cognitive capacity may lead to accidents and injuries. The ability to monitor changes in cognitive capacity over time would have benefits for planning and decision making in both civilian and military settings, and may inform interventions and performance-enhancement techniques.

A frequently used method for assessing cognitive capacity in real time is the psychomotor vigilance task (PVT), which is a test that measures reaction time and attentiveness, developed primarily as an in-laboratory assessment. Despite recent efforts to integrate PVT into portable and/or wearable devices, the test itself is still largely inappropriate for operational environments because it requires the tested individuals to disengage from their primary task and spend between two and 20 minutes participating in the PVT. Some success has been achieved in developing nonobtrusive assessment methods that are based on ocular behavior, such as the percentage of eyelid closure (e.g., PERCLOS) and ocular dynamics (e.g., SmoothEye). These methods hold promise for monitoring cognitive capacity in field settings but nonetheless require specialized equipment and/or specific environmental conditions to function accurately.

VOICE CHARACTERIZATION AND FEATURE SELECTION

In ongoing work with the U.S. Army and U.S. Air Force, we have been investigating voice-based technologies for real-time assessment of cognitive capacity. We are using a variety of standard acoustic speech features in combination with the timing and coordination features developed by Lincoln Laboratory. Voice-based assessments for cognitive capacity are attractive because they can be made nonobtrusive, can be implemented in hardware that uses conventional microphones, and have the potential to be adapted to standard-issue communications platforms used in operational environments, such as airplane cockpits. Early development and validation have been taking place on data collected at Lincoln Laboratory. Data collections with U.S. Army and U.S. Air Force partners are ongoing and will substantially advance early results and validation efforts.
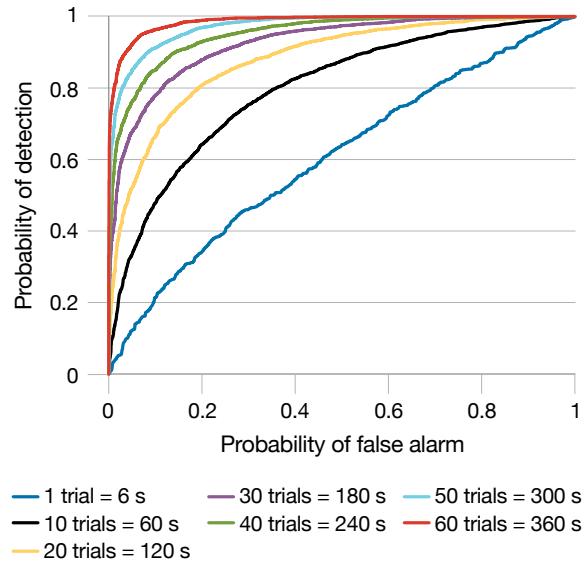
(a)
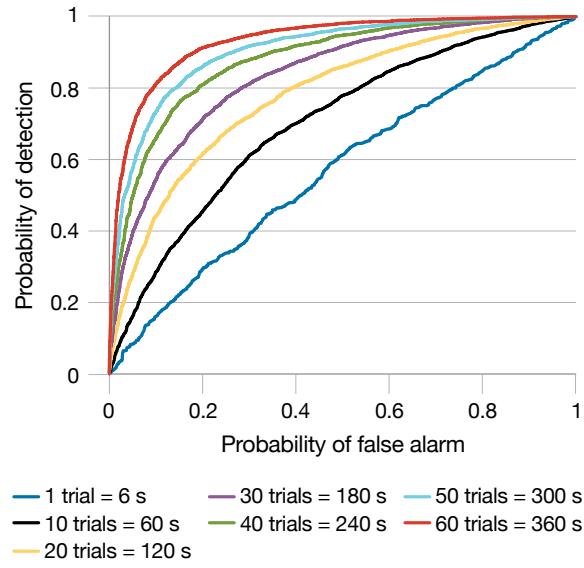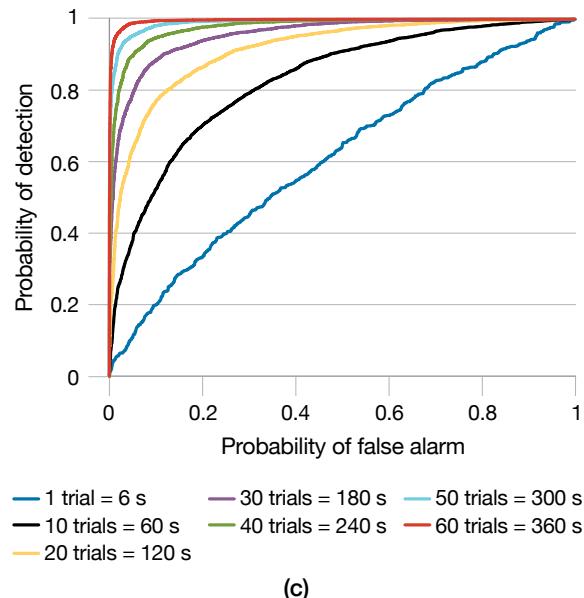
EXAMPLE PREDICTION

We conducted a preliminary proof of concept of the biomarkers pipeline in the domain of cognitive fatigue by applying it to data collected at Lincoln Laboratory. We chose the longitudinal estimation of typical, daily fluctuations of fatigue for a single individual. This single subject was recorded reading the Caterpillar passage [80] three times each day (at approximately 9 a.m., 12 p.m., and 4 p.m.) for five consecutive days (Monday through Friday of a normal work week). At these times, the subject also performed a 10-minute PVT session to assess level of fatigue. Mean reaction time was used to quantify the level of fatigue. A suite of the Laboratory's low- and high-level features were extracted from the speech audio and used as a basis for quantifying fatigue levels. Principal components analysis was performed on the audio features to achieve dimensionality reduction, keeping the first five most prominent components. A linear model was trained with a leave-one-session-out cross validation to predict mean reaction time from the audio features. Results in Figure 19 show a strong correlation ($r^2 = 0.64$, $p < 0.01$) between actual and estimated reaction time, with a root-mean-square error (RMSE) of 9.6 milliseconds, or approximately 15.5 percent of the observed range of mean reaction times in the dataset. These results, though preliminary, demonstrate the promise of vocal biomarkers for cognitive fatigue assessment and, more generally, for cognitive status monitoring. The ability to

(b)

(c)

FIGURE 18. The fused probability of detection versus false alarm for the modalities of EEG (a), audio and video (b), and EEG, audio, and video (c) is compared. Each panel gives ROCs as a function of increasing number of trials from 1 to 60, corresponding to 6 seconds to 360 seconds (6 minutes) for low and high cognitive loads.

predict objective measures of fatigue with an estimation error of approximately 16 percent may be relevant and actionable for individuals on sleep-restricted schedules during which fatigue levels have been shown to vary by 35 percent or more [81]. Preliminary results on larger datasets have revealed improved performance compared

to this modest-sized proof of concept. Our expanded effort in this area is also investigating the independent and joint effects of cognitive load and fatigue conditions in voice features, showing the changing discriminatory ability of cognitive load under different cognitive fatigue conditions [82].

Other Modalities

Lincoln Laboratory is also expanding to other behavioral measures, including fine motor (e.g., eye tracking and hand dexterity) and gross motor movements (e.g., static and dynamic balance), physiological measures (e.g., heart rate and skin conductance and temperature), and brain-computer interfaces (e.g., EEG measures), along with corresponding neurobiophysical models, for detecting, phenotyping, and monitoring brain trauma and disease. For example, eye tracking is included in our high school athlete data collection for preclinical mTBI in our collaboration with Purdue University and also in a cognitive load data collection in collaboration with the Combat Capabilities Development Command Soldier Center (CCDC Soldier Center), formerly the U.S. Army Natick Soldier Research, Development, and Engineering Center. In each case, eye-tracking features have contributed significantly to the prediction of cognitive decline [75] and cognitive load level [83], respectively. Physiological measures also play a role in our past and ongoing investigations of depression (as part of our multimodal study with the Wyss Institute at Harvard Medical School) and of cognitive overload and fatigue (as part of our multimodal studies with USARIEM and CCDC Soldier Center). In addition to using EEG analysis in cognitive load level prediction, we are using it to obtain insights into the importance of active brain regions during visual and auditory memory tasks [83]. Finally, in this same context of cognitive load prediction, we are investigating the effect of movement in cognition. One example involves recording speech, EEG (64-channel Neuroscan system), and full-body motion capture (17-accelerometer Xsens System) from 10 subjects in both standing and walking conditions while they are engaged in an auditory working memory task. Here, we are exploring the effect of cognitive load on timing and coordination features within and across the three modalities, their interrelationships, and the consequence of multitasking [84].
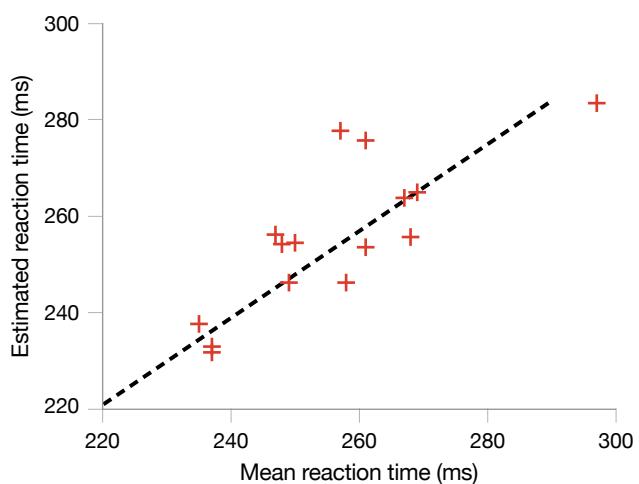
FIGURE 19. Preliminary results of using speech-based biomarkers for predicting reaction time are plotted. Reaction time, as measured with the psychomotor vigilance task, is sometimes taken as a gold standard for assessment of alertness and cognitive fatigue. A linear model was trained with a leave-one-session-out cross validation to predict reaction time from speech features, using data collected at Lincoln Laboratory of a single subject assessed three times daily for one work week.

Providing sensitive detection simultaneously with specific phenotyping is an inherent challenge facing the assessment of many different neurological conditions. An especially prevalent example of these challenges can be seen in the domain of mTBI, where a wide variety of sensorimotor (e.g., vestibular deficits) and cognitive (e.g., attention deficits) impairments lead to subtle and heterogeneous signs and symptoms. Clinical assessments in this domain tend to be subjective and insensitive, and neural imaging is expected to be negative. Moreover, diagnosis is typically focused on classifying the injury into broad categories: mild, moderate, or severe. There is a pressing need for technologies that can provide individualized physiological phenotypes in order to facilitate diagnostic specificity, which is often complicated by the co-occurrence of conditions with similar symptom profiles (e.g., PTSD and depression), as well as to tailor treatment approaches to individual needs and more accurately monitor risk and post-injury recovery.

In one program, Lincoln Laboratory staff—in collaboration with clinical and domain experts from Spaulding Rehabilitation Hospital, Massachusetts General Hospital, and Harvard Medical School—are developing

(a)
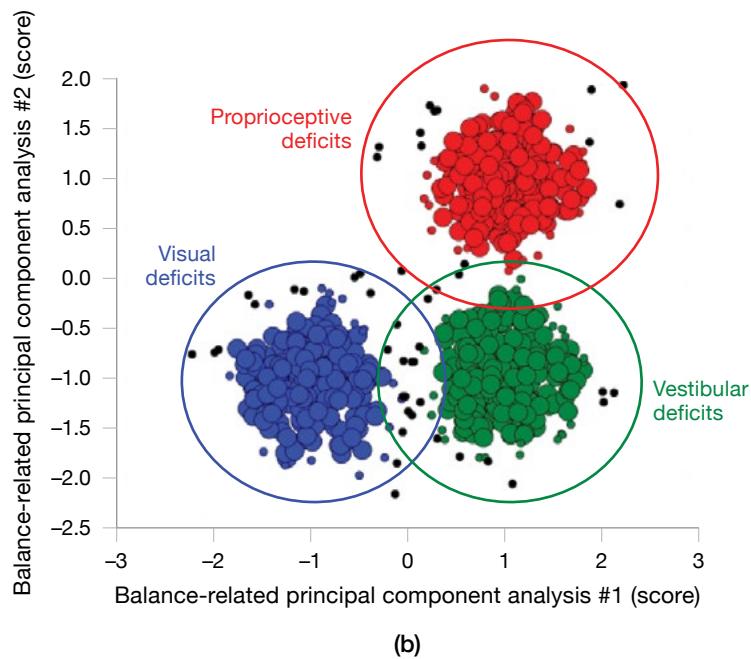
(b)

FIGURE 20. Neurocomputational model–based approaches to phenotyping mTBI are being developed in the virtual reality environment at the Laboratory's Sensorimotor Technology Realization in Immersive Virtual Environments (STRIVE) Center (a). Behavioral and physiological responses to sensorimotor provocative tests are measured, driving neurocomputational models of motor control in normal and impaired individuals. This testing can facilitate system identification of impairment in neurological control mechanisms involving vestibular, visual, and proprioceptive sensory feedback (b).

methods for individualized phenotyping of mTBI-related impairments through the analysis of balance and gait. Balance and gait show frequent impairment in mTBI from a variety of causes (e.g., vestibular, visual-motor, and/or proprioceptive sensory feedback impairment) and therefore hold promise for enabling phenotyping. Two parallel approaches are underway (see Figure 20).

The first is the development of sensorimotor provocative tests, conducted on a flexible treadmill platform in the virtual reality environment at the Laboratory's Sensorimotor Technology Realization in Immersive Virtual Environments (STRIVE) Center. These tests are designed to bring out latent impairments not observable in activities of daily living. The second approach

is the development of neurocomputational models of motor control in normal and impaired individuals to facilitate system identification of neurological control mechanisms. These approaches are synergistic, with data from STRIVE Center experiments used to refine control models and models used to inform the design of sensorimotor provocative tests.

Long-Term Vision

In the Laboratory's Human Health and Performance Systems Group, we have previous or ongoing projects in detecting and tracking amyotrophic lateral sclerosis (ALS) [84], dementia [86], and autism spectrum disorder [87]. A common theme to detecting and tracking these conditions is the discovery of behavioral-based biomarkers that reflect a change or decline in brain functioning as manifested in motor control, and more specifically changes in timing and coordination within and across components of behaviors. It is hypothesized that these relations are associated with neural coordination across different parts of the brain that are essential in motor control.

As a final application, we mention our introduction into an area of immediate urgency: detecting and tracking COVID-19 (the novel coronavirus) through asymptomatic and symptomatic stages. Given the physiologically based insult to breathing functions [88] and the growing evidence of neurological deficits present in COVID-19 [89, 90], we hypothesize that biomarkers derived from measures of vocal-subsystem coordination that includes both lower and upper respiratory systems may provide a sensitive indicator of COVID-19, most importantly in its asymptomatic stages. Preliminary results with audio interviews of two subjects reveal strong effect sizes in distinguishing pre-COVID-19 (pre-exposure) from post-COVID-19 (after positive diagnosis but still asymptomatic) by using intensity of breathing (respiration during speech), coordination of respiration and pitch (fundamental frequency), and coordination of pitch and articulatory motion. Morphology of eigenvalue Cohen's d effect sizes indicates a constricted breathing and reduced complexity of coordinated subsystem movement. Although preliminary results are promising, we clearly need to validate these results and address confounding influences with larger, more controlled datasets.

With our timing and coordination-based features, we have achieved effective detection across a variety of application areas. Nevertheless, to more strongly validate performance in all areas, we need to move from the lab and clinic to the field. To provide access to larger populations, we are currently translating our data collections and detection algorithms to apps using mobile technology, e.g., smartphones and tablets. In the area of depression, we are working with the commercial entity Sonde Health; in traumatic brain injury, we are collaborating with the U.S. Army Medical Materiel Development Activity (USAMMDA) to develop a mobile device approved by the U.S. Food and Drug Administration; and in cognitive load and fatigue, we are collaborating with USARIEM toward a predictive algorithm for assessment within a technology transfer agreement. Toward our goal of early warning and tracking of COVID-19, it will be essential to address potential confounders, such as different recording environments and channels, unbalanced data quantities, and changes in underlying vocal status from pre-COVID-19 exposure to post-COVID-19 diagnosis. Finally, it will be essential to understand the specificity of our proposed biomarkers; for example, these proposed biomarkers must be able to differentiate COVID-19 from the typical flu and flu-like conditions resulting in various forms of inflammation.

A growing path in our biomarker research and development involves pairing our behavioral approaches with neurocomputational biophysical modeling and with clinical observation. This trifold strategic approach fuses empirical measurements, phenomenological (data-driven) models, and mechanistic models.

Our long-term vision is to move away from predictions based on group analysis and to use the trifold fusion to improve assessments of mechanisms underlying individual impairments, leading to personalized design of patient care monitoring and intervention systems. Our ultimate objective is to develop patient-specific neurocomputational models of human behavior that integrate sensorimotor impairments for early warning and improved specificity and phenotyping.

References

1. World Health Organization, Mental Health, https://www.who.int/mental_health/en/.
2. Office of Suicide Prevention, "Suicide Among Veterans and Other Americans," U.S. Department of Veterans Affairs, 3 Aug. 2016, available at http://www.mentalhealth.va.gov/docs/2016suicidedatareport.pdf.
3. World Health Organization, "Dementia Fact Sheet," available at http://www.who.int/en/news-room/fact-sheets/detail/dementia.
4. Centers for Disease Control and Prevention, "Community Report on Autism 2018," U.S. Department of Health and Human Services.
5. T. Elianan and L.H. Joycox, eds., *Invisible Wounds of War: Psychological and Cognitive Injuries, Their Consequences, and Services to Assist Recovery*. Santa Monica: RAND Corporation, 2008.
6. S.E. Lively, D.B. Pisoni, W. Van Summers, and R.H. Bernacki, "Effects of Cognitive Workload on Speech Production: Acoustic Analyses and Perceptual Consequences," *Journal of the Acoustical Society of America*, vol. 93, no. 5, 1993, pp. 2962–2973.
7. B. Yin, F. Chen, N. Ruiz, and E. Ambikairajah, "Speech-Based Cognitive Load Monitoring System," paper in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2008, pp. 2041–2044.
8. B. Yin and F. Chen, "Towards Automatic Cognitive Load Measurement from Speech Analysis," paper in *Proceedings of the 12th International Human-Computer Interaction Conference, Interaction Design and Usability*, 2007, pp. 1011–1020.
9. M.A. Khawaja, N. Ruiz, and F. Cheng, "Think Before You Talk: An Empirical Study of Relationship between Speech Pauses and Cognitive Load," paper in *Proceedings of the 20th Australasian Computer-Human Interaction Conference*, 2008, pp. 335–338.
10. P. Le, J. Epps, H.C. Choi, and E. Ambikairajah, "A Study of Voice Source- and Vocal Tract-Based Features in Cognitive Load Classification," paper in *Proceedings of the International Conference on Pattern Recognition*, 2010, pp. 4516–4519.
11. P. Zarjam, J. Epps, and F. Chen, "Characterizing Working Memory Load Using EEG Delta Activity," paper in *Proceedings of the 19th European Signal Processing Conference*, 2011, pp. 1554–1558.
12. P. Zarjam, J. Epps, and F. Chen, "Evaluation of Working Memory Load Using EEG Signals," paper in *Proceedings*
13. N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T.F. Quatieri, "A Review of Depression and Suicide Risk Assessment using Speech Analysis," *Speech Communication*, vol. 17, 2015, pp. 10–49.
14. E. Moore, M. Clements, J. Peifer, and L. Weisser, "Analysis of Prosodic Variation in Speech for Clinical Depression," paper in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2003, pp. 2925–2928.
15. J.C. Mundt, P.J. Snyder, M.S. Cannizzaro, K. Chappie, and D.S. Geralts, "Voice Acoustic Measures of Depression Severity and Treatment Response Collected via Interactive Voice Response (IVR) Technology," *Journal of Neurolinguistics*, vol. 20, no. 1, 2007, pp. 50–64.
16. A. Ozdas, R.G. Shiavi, S.E. Silverman, M.K. Silverman, and D.M. Wilkes, "Investigation of Vocal Jitter and Glottal Flow Spectrum as Possible Cues for Depression and Near-Term Suicidal Risk," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 9, 2004, pp. 1530–1540.
17. J.K. Darby, N. Simmons, and P.A. Berger, "Speech and Voice Parameters of Depression: A Pilot Study," *Journal of Communication Disorders*, vol. 17, no. 2, 1984, pp. 75–85.
18. J.F. Greden and B.J. Carroll, "Psychomotor Function in Affective Disorders: An Overview of New Monitoring Techniques," *The American Journal of Psychiatry*, vol. 138, no. 11, 1981, pp. 1441–1448.
19. J.R. Orozco-Arroyave, J.D. Arias-Londoño, J.F. Vargas-Bonilla, M.C. González-Rátiva, and E. Nöth, "New Spanish Speech Corpus Database for the Analysis of People Suffering from Parkinson's Disease," paper in *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, 2014, pp 342–347.
20. P. Wang, F. Barrett, E. Martin, M. Milonova, R.E. Gur, R.C. Gur, et al., "Automated Video-Based Facial Expression Analysis of Neuropsychiatric Disorders," *Journal of Neuroscience Methods*, vol. 168, no. 1, 2008, pp. 224–238.
21. M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, et al., "AVEC 2014: 3D Dimensional Affect and Depression Recognition Challenge," paper in *Proceedings of the Fourth International Workshop on Audio/Visual Emotion Challenge*, 2013, pp. 3–10.
22. W. Gaebel and W. Wölwer, "Facial Expression and Emotional Face Recognition in Schizophrenia and Depression," *European Archives of Psychiatry and Clinical Neuroscience*, vol. 242, no. 1, 1992, pp. 46–52.
23. G.E. Schwartz, P.L. Fair, P. Salt, M.R. Mandel, and G.L. Klerman, "Facial Expression and Imagery in Depression: An Electromyographic Study," *Psychosomatic Medicine*, vol. 38, no. 5, 1976, pp. 337–347.
24. P. Ekman, W.V. Freisen, and S. Ancoli, "Facial Signs of Emotional Experience," *Journal of Personality and Social Psychology*, vol. 39, no. 6, 1980, pp. 1125–1134.

25. P. Zarjam, J. Epps, and N.H. Lovell, "Beyond Subjective Self-Rating: EEG Signal Classification of Cognitive Workload," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 4, 2015, pp. 301–310.

26. M. Fava and K.S Kendler, "Major Depressive Disorder," *Journal of Neuroscience*, vol. 28, no. 2, 2000, pp. 335–341.

27. G. Littlewort, J. Whitehill, T. Wu, I.R. Fasel, M. Frank, J.R. Movellan, and M.S. Bartlett, "The Computer Expression Recognition Toolbox (CERT)," paper presented at the Ninth IEEE International Conference on Automatic Face and Gesture Recognition, Santa Barbara, 21–23 Mar. 2011.

28. D.D. Mehta, D. Rudoy, and P.J. Wolfe, "Kalman-Based Autoregressive Moving Average Modeling and Inference for Formant and Antiformant Tracking," *The Journal of the Acoustical Society of America*, vol. 132, no. 3, 2012, pp. 1732–46.

29. A.C. Trevino, T.F. Quatieri, and N. Malyska, "Phonologically-Based Biomarkers for Major Depressive Disorder," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, 2011, pp. 1–18.

30. G.J. Canter, "Speech Characteristics of Patients with Parkinson's Disease: I. Intensity, Pitch, and Duration," *Journal of Speech and Hearing Disorders*, vol. 28, no. 3, 1963, pp. 221–229.

31. J.A. Logemann, H.B. Fisher, B. Boshes, and E.R. Blonsky, "Frequency and Cooccurrence of Vocal Tract Dysfunctions in the Speech of a Large Sample of Parkinson Patients," *Journal of Speech and Hearing Disorders*, vol. 43, no. 1, 1978, pp. 47–57.

32. J. E. Sussman and K. Tjaden, "Perceptual Measures of Speech from Individuals with Parkinson's Disease and Multiple Sclerosis: Intelligibility and Beyond," *Journal of Speech, Language, and Hearing Research*, vol. 55, no. 4, 2012, pp. 1208–1219.

33. S. Skodda and U. Schlegel, "Speech Rate and Rhythm in Parkinson's Disease," *Movement Disorders*, vol. 23, no. 7, 2008, pp. 985–992.

34. J. Rusz, R. Cmejla, T. Tykalova, H. Ruzickova, J. Klempir, V. Majerova, et al., "Imprecise Vowel Articulation as a Potential Early Marker of Parkinson's Disease: Effect of Speaking Task," *The Journal of Acoustical Society of America*, vol. 134, no. 3, 2013, pp. 2171–2181.

35. C.G. Goetz, B.C. Tilley, S.R. Shaftman, G.T. Stebbins, S. Fahn, P. Martinez-Martin, et al., "Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale Presentation and Clinimetric Testing Results," *Movement Disorders*, vol. 23, no. 15, 2008, pp. 2129–2170.

36. M. Falcone, N. Yadav, C. Poellabauer, and P. Flynn, "Using Isolated Vowel Sounds for Classification of Mild Traumatic Brain Injury," paper in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2013, pp. 7577–7581.

37. T.M. Talavage, E.A. Nauman, E.L. Breedlove, U. Yoruk, A.E. Dye, K.E. Morigaki, et al., "Functionally-Detected Cognitive Impairment in High School Football Players without Clinically-Diagnosed Concussion," *Journal of Neurotrauma*, vol. 31, no. 4, 2014, pp. 327–338.

38. C. Cao, R.L. Tutwiler, and S. Slobounov, "Automatic Classification of Athletes with Residual Functional Deficits Following Concussion by Means of EEG Signal Using Support Vector Machine," *IEEE Transactions of Neural Systems and Rehabilitation and Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, vol. 16, no. 4, 2008, pp. 327–335.

39. C. Poellabauer, N. Yadav, L. Daudet, S.L. Schneider, C.B. Busso, and P. Flynn, "Challenges in Concussion Detection Using Vocal Acoustic Biomarker," *IEEE Access*, vol. 3, 2015, pp. 1143–1160.

40. A.L. Russo, J. Verghese, A.L. Metti, R.M. Boudreau, H.J. Aizenstein, S. Kritchevsky, et al., "Slowing Gait and Risk for Cognitive Impairment: The Hippocampus as a Shared Neural Substrate," *Neurology*, vol. 89, no. 4, 2017, pp. 336–342.

41. U.U. Samadani, R. Ritlop, M. Reyes, E. Nehrbass, M. Li, E. Lamm, et al., "Eye Tracking Detects Disconjugate Eye Movements Associated with Structural Traumatic Brain Injury and Concussion," *Journal of Neurotrauma*, vol. 32, 2015, pp. 548–556.

42. E. Gowen and A. Hamilton, "Motor Abilities in Autism: A Review Using a Computational Context," *Journal of Autism and Development Disorders*, vol. 43, no. 2, 2013, pp. 323–344.

43. T.F. Quatieri, J.R. Williamson, C.J. Smalt, J. Perricone, T. Patel, L.J. Brattain, et al., "Multi-modal Biomarkers to Discriminate Cognitive State," *The Role of Technology in Clinical Neuropsychology*. Eds. R.L Kane and T.D. Parsons. New York: Oxford University Press, 2017.

44. B. Yu, T.F. Quatieri, J.W. Williamson, and J. Mundt, "Prediction of Cognitive Performance in an Animal Fluency Task Based on Rate and Articulatory Markers," paper in *Proceedings of the 15th Annual Conference of the International Speech Communication Association*, 2014, pp. 1038–1042.

45. B. Yu, T.F. Quatieri, J.R. Williamson, and J.C. Mundt, "Cognitive Impairment Prediction in the Elderly Based on Vocal Biomarkers," paper in *Proceedings of the 16th Annual Conference of the International Speech Communication Association*, 2015.

46. J.R. Williamson, T.F. Quatieri, B.S. Helfer, R. Horwitz, B. Yu, and D.D. Mehta, "Vocal Biomarkers of Depression Based on Motor Incoordination," paper in *Proceedings of the Fourth International Workshop on Audio/Visual Emotion Challenge*, 2013, pp. 65–72.

47. T.F. Quatieri, J.R. Williamson, C.J. Smalt, T. Patel, J. Perricone, D.D. Mehta, et al., "Vocal Biomarkers to Discriminate Cognitive Load in a Working Memory Task," paper in *Proceedings of the 16th Annual Conference of the International Speech Communication Association*, 2015, pp. 2684–2688.

48. F.H. Guenther, *Neural Control of Speech*. Cambridge, Mass.: MIT Press, 2018.

49. K.A. Gola, Tal Shany-Ur, P. Pressman, I. Sulman, E. Galeana, H. Paulsen, et al., "A Neural Network Underlying Intentional Emotional Facial Expression in Neurodegenerative Disease," *NeuroImage: Clinical*, vol. 14, 2017, pp. 672–678.

50. T.F. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice*. Upper Saddle River, N.J.: Prentice Hall, 2002.

51. R.L. Rabiner and B.-H.Juang, *Fundamentals of Speech Recognition*. Upper Saddle River, N.J.: Prentice Hall, 1993.

52. P. Ekman and E.L. Rosenberg, eds., *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, 2nd ed. New York: Oxford University Press, 2005.

53. T.F. Quatieri and N. Malyska, "Vocal-Source Biomarkers for Depression: A Link to Psychomotor Activity," paper in *Proceedings of the 13th Annual Conference of the International Speech Communication Association*, 2012, pp. 1059–1062.

54. J.R. Williamson, T.F. Quatieri, B.S. Helfer, G. Ciccarelli, and D.D. Mehta, "Vocal and Facial Biomarkers of Depression Based on Motor Incoordination and Timing," paper in *Proceedings of the Fourth International Workshop on Audio/Visual Emotion Challenge*, 2014, pp. 65–72.

55. B.S Helfer, T.F. Quatieri, J.R. Williamson, L. Keyes, B. Evans, W.N. Greene, et al., "Articulatory Dynamics and Coordination in Classifying Cognitive Change with Preclinical mTBI," paper in *Proceedings of the 15th Annual Conference of the International Speech Communication Association*, 2014.

56. A. Trevino, T.F. Quatieri, and N. Malyska, "Phonologically-Based Biomarkers for Major Depressive Disorder," *EURASIP Journal on Advances in Signal Processing: Special Issue on Emotion and Mental State Recognition from Speech*, vol. 42, no. 1, 2011, pp. 2011–2042.

57. J.R. Williamson, T.F. Quatieri, B.S. Helfer, J. Perricone, S.S. Ghosh, G. Ciccarelli, and D.D. Mehta, "Segment-Dependent Dynamics in Predicting Parkinson's Disease," paper in *Proceedings of the 16th Annual Conference of the International Speech Communication Association*, 2015, pp. 518–522.

58. J.R. Williamson, D.W. Bliss, and D.W. Browne, "Epileptic Seizure Prediction Using the Spatiotemporal Correlation Structure of Intracranial EEG," paper in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011, pp. 665–668.

59. J.R. Williamson, D.W. Bliss, D.W. Browne, and J.T. Narayanan, "Seizure Prediction Using EEG Spatiotemporal Correlation Structure," *Epilepsy and Behavior*, vol. 25, no. 2, 2012, pp. 230–238.

60. R. Horwitz, T.F. Quatieri, B.S. Helfer, B. Yu, J.R. Williamson, and J. Mundt, "On the Relative Importance of Vocal Source, System, and Prosody in Human Depression," paper in *Proceedings of the 2013 IEEE International Conference on Body Sensor Networks*, 2013, pp. 1–6.

61. P.J. Jackson and C.H. Shadle, "Performance of the Pitch-Scaled Harmonic Filter and Applications in Speech Analysis," paper in *Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2000, pp. 1311–1314.

62. P.J. Jackson and C.H. Shadle, "Pitch-Scaled Estimation of Simultaneous Voiced and Turbulence-Noise Components in Speech," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 7, 2011, pp. 713–726.

63. J. Kane, T. Drugman, and C. Gobl, "Improved Automatic Detection of Creak," *Computer Speech and Language*, vol. 27, no. 4, 2013, pp. 1028–1047.

64. T. Drugman, Toolbox, http://tcts.fpms.ac.be/~drugman/Toolbox/.

65. D.A. Reynolds, T.F. Quatieri, and R.B. Dunn, "Speaker Verification using Adapted Gaussian Mixture Models," *Digital Signal Processing*, vol. 10, no. 1, 2000, pp. 19–41.

66. W. Shen, C. White, and T.J. Hazen, "A Comparison of Query-by-Example Methods for Spoken Term Detection," paper in *Proceedings of the 10th Annual IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2009, pp. 2143–2146.

67. B.S. Helfer, J.R. Williamson, B.A. Miller, J. Perricone, and T.F. Quatieri, "Assessing Functional Neural Connectivity as an Indicator of Cognitive Performance," paper presented at the 5th Workshop on Machine Learning and Interpretation in Neuroimaging, Montreal, 11–12 Dec. 2015.

68. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. Washington, D.C.: American Psychiatric Association Publishing, 2000.

69. P. Boersma, "Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound," paper in *Proceedings of the Institute of Phonetic Sciences*, vol. 17, 1993, pp. 97–110.

70. G.J. Canter, "Speech Characteristics of Patients with Parkinson's Disease: II. Physiological Support for Speech," *Journal of Speech and Hearing Disorders*, vol. 30, no. 1, 1965, pp. 44–49.

71. G.J. Canter, "Speech Characteristics of Patients with Parkinson's Disease: III. Articulation, Diadochokinesis, and Over-All Speech Adequacy," *Journal of Speech and Hearing Disorders*, vol. 30, no. 3, 1965, pp. 217–224.

72. J.A. Logemann and H.B. Fisher, "Vocal Tract Control in Parkinson's Disease," *Journal of Speech and Hearing Disorders*, vol. 46, no. 4, 1981, pp. 348–352.

73. S. Skodda, "Aspects of Speech Rate and Regularity in Parkinson's Disease," *Journal of Neurological Sciences*, vol. 310, no. 1–2, 2011, pp. 231–236.

74. R. Netsell and D. Lefkowitz, "Speech Production Following Traumatic Brain Injury: Clinical and Research Implications," *ASHA Neurophysiology and Neurogenic Speech and Language Disorders*, vol. 2, no. 4, pp. 1–8.

75. B.S. Helfer, T.F. Quatieri, L. Keyes, B. Evans, W.N. Greene, T. Vian, et al., "Voice and Ocular Dynamics in Classifying Cognitive Change with Preclinical mTBI," poster presented at the Fourth Annual Traumatic Brain Injury Conference, Washington, D.C., 16–17 Apr. 2014.

76. ImPACT Applications Inc., "Immediate Post-Concussion Assessment Testing (ImPACT®) Test Technical Manual Technical Manual and Psychometric Data," available at impacttest.com, 2013.

77. A.C. Lammert, J.R. Williamson, A. Hess, T. Patel, T.F. Quatieri, H.J. Liao, et al., "Noninvasive Estimation of Cognitive Status in Mild Traumatic Brain Injury Using Speech Production and Facial Expression," paper in *Proceedings of the Seventh International Conference on Affective Computing and Intelligent Interaction*, 2017, pp. 105–110.

78. K. Watanabe and S. Funahashi, "Neural Mechanisms of Dual-Task Interference and Cognitive Capacity Limitation in the Prefrontal Cortex," *Nature Neuroscience*, vol. 17, 2014, pp. 601–611.

79. P. Zarjam, J. Epps, and F. Che, "Spectral EEG Features for Evaluating Cognitive Load," paper in *Proceedings of the 33rd Annual International Conference of IEEE Engineering in Medicine and Biology Society*, 2011, pp. 3841–3844.

80. R. Patel, K. Connaghan, D. Franco, E. Edsall, D. Forgit, L. Olsen, et al., "'The Caterpillar': A Novel Reading Passage for Assessment of Motor Speech Disorders," *American Journal of Speech-Language Pathology*, vol. 22, no. 1, 2013, pp. 1–9.

81. D.P. Redmond, M.L. Johnson, D.R. Thorne, G. Belenky, T.J. Balkin, W.F. Storm, et al., "Fatigue Models for Applied Research in Warfighting," *Aviation, Space, and Environmental Medicine*, vol. 75, no. 3, 2004, pp. 44–60.

82. J. Sloboda, A. Lammert, J. Williamson, C. Smalt, D.D. Mehta, I. Curry, et al., "Vocal Biomarkers for Cognitive Performance Estimation in a Working Memory Task," paper in *Proceedings of the 19th Annual Conference of the International Speech Communication Association*, 2018.

83. M.A. Nolan, J.R. Williamson, L.J. Brattain, M.D. Eddy, J.M. Moran, C.J. Smalt, and T.F. Quatieri, "Predicting Variations in Cognitive Load: A Multimodal Approach," poster presented at Neuroscience 2016, San Diego, 12–16 Nov. 2016.

84. T.F. Quatieri, B. Yu, J. Perricone, A. Lammert, M. Nolan, D. Mehta, and J.S. Palmer, "Energy and Correlation Analysis of Gait and EEG Representations During an Auditory Working Memory Task, Movement and Cognition," poster presented at the Movement and Cognition International Conference at Harvard University School of Medicine, Boston, 27–29 Jul. 2018.

85. R. Horwitz-Martin, T. Quatieri, A. Lammert, J. Williamson, Y. Yunusova, E. Godoy, et al., "Relation of Automatically Extracted Formant Trajectories with Intelligibility Loss and Speaking Rate Decline in Amyotrophic Lateral Sclerosis," paper in *Proceedings of the 18th Annual Conference of the International Speech Communication Association*, 2017.

86. B. Yu, J.R. Williamson, J.C. Mundt, and T.F. Quatieri, "Speech-Based Automated Cognitive Impairment Detection from Remotely-Collected Cognitive Test Audio," *IEEE Access*, vol. 6, no. 1, 2018, pp. 40494–40505.

87. T.F. Quatieri, J. O'Rourke, L. Nowinski, D. Hannon, A. Lammert, J. Williamson, et al., "Quantifying Fine Motor Dependencies in Autism Spectrum Disorder," *Neuroscience*, 2018.

88. World Health Organization, "Clinical Management of Severe Acute Respiratory Infection (SARI) When COVID-19 Disease is Suspected: Interim Guidance Report," 13 Mar. 2020, available at https://www.who.int/publications-detail/clinical-management-of-severe-acute-respiratory-infection-when-novel-coronavirus-(ncov)-infection-is-suspected.

89. L. Mao, M. Wang, S. Chen, Q. He, J. Chang, C. Hong, et al., "Neurological Manifestations of Hospitalized Patients with COVID-19 in Wuhan, China: A Retrospective Case Series Study," *JAMA Neurology*, published online 10 Apr. 2020.

90. C.H. Yan, F. Faraji, D.P. Prajapati, C.E. Boone, and A.S. DeConde, "Association of Chemosensory Dysfunction and Covid-19 in Patients Presenting with Influenza-like Symptoms," *International Forum of Allergy & Rhinology*, published online 12 Apr. 2020.

About the Authors

Thomas F. Quatieri, a senior member of the technical staff in the Human Health and Performance Systems Group, is involved in bridging human language and bioengineering research and technologies. Within this group, he has initiated and developed major R&D and technology transition programs in speech and auditory signal processing and neuro-biophysical modeling with application to detection and monitoring of neurological, neurotraumatic, and stress conditions. He has been an author on more than 200 publications, holds 11 patents, and authored the textbook *Discrete-Time Speech Signal Processing: Principles and Practice*. He also holds a faculty appointment in the Harvard-MIT Speech and Hearing Bioscience and Technology Program. He developed the MIT graduate course Digital Speech Processing and is active in advising graduate students on the MIT and Harvard campuses. He is a recipient of four IEEE Transactions Best Paper Awards and the 2010 MIT Lincoln Laboratory Best Paper Award. He led the Lincoln Laboratory team that won the 2013 and 2014 AVEC Depression Challenges and the 2015 MIT Lincoln Laboratory Team Award for their work on vocal and facial biomarkers. He has served on the IEEE Digital Signal Processing Technical Committee, the IEEE Speech and Language Technical Committee, and the IEEE James L. Flanagan Speech and Audio Awards Committee. He has served on National Institutes of Health and National Science Foundation panels, been an associate editor for the *IEEE Transactions on Signal Processing*, and is an associate editor of *Computer, Speech, and Language*. He has been an invited speaker at conferences, workshops, and meetings, most recently as a keynote speaker at the 2019 Speech Science and Technology Conference. He is a Fellow of the IEEE and a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, International Speech and Communication Association (ICSA), Society for Neuroscience, Association for Research in Otolaryngology (ARO), and the

THOMAS F. QUATIERI, JAMES R. WILLIAMSON, ADAM C. LAMMERT, KRISTIN J. HEATON, AND JEFFREY S. PALMER

Acoustical Society of America (ASA). He holds a bachelor's degree (summa cum laude) from Tufts University, and master's, electrical engineering, and doctoral degrees from MIT.

James R. Williamson is a member of the technical staff in the Human Health and Performance Systems Group. His work focuses on detecting and modeling neurocognitive and physical state through the analysis of speech, facial expressions, gait, and body-sensing modalities. Prior to joining Lincoln Laboratory, he earned a bachelor's degree in psychology from the University of Massachusetts Amherst and a doctoral degree in cognitive and neural systems from Boston University. He also served as a research assistant professor at Boston University.

Adam C. Lammert was a member of the technical staff in the Bioengineering Systems and Technologies Group beginning in 2015, primarily focusing on neurocognitive analysis, motor control modeling, and speech signal processing. He currently holds the position of assistant professor at Worchester Polytechnical Institute. From 2006 to 2008, he was lab manager of speech and hearing research at the Veterans Affairs Northern California Health Care System in Martinez, California. He earned a doctoral degree in computer science from the University of Southern California, where he was a member of the Signal Analysis and Interpretation Laboratory and the Hearing and Communication Neuroscience Program at the House Research Institute. He was awarded the Raymond H. Stetson Scholarship in Phonetics and Speech Science by the Acoustical Society of America in 2013. Prior to joining Lincoln Laboratory, he was a visiting assistant professor in the Computer Science Department at Swarthmore College. He has published more than 45 papers, several of which have been awarded special status. He earned a bachelor's degree in cognitive science from Vassar College and a master's degree in computer science from North Carolina State University.

Kristin Heaton joined the Military Performance Division of the U.S. Army Research Institute of Environmental Medicine in Natick, Massachusetts, as a research psychologist in 2005. In July 2019, she took on the role of deputy chief in the Military Performance Division and currently holds the position of supervisory research psychologist. She earned a doctoral degree in psychology from the University of Maryland, College Park in 2002. Specializing in the field of neuropsychology, she completed her postdoctoral clinical fellowship in neuropsychology with the Boston Consortium in Clinical Psychology, a joint program

of the VA Boston Healthcare System, Boston University, and Harvard University. Previously, she was an adjunct research assistant professor at Boston University School of Public Health. Her research focuses broadly on the development, application, and validation of tools and techniques for assessment of neuro-behavioral health and performance in military operational and training environments. Her current research portfolio includes studies examining the following: unobtrusive, naturally occurring behavioral markers (e.g., speech, facial expression, and physiological responses) for prediction and monitoring of warfighter performance (cognitive, mental, physical); the impacts of military occupational, environmental, and neurotoxicant (fuel products, pesticides) exposures on warfighter neurological health and performance outcomes; and novel approaches such as electrical brain stimulation for optimizing and enhancing warfighter physical, mental, and cognitive performance. She is currently an active member of several federal agency (Department of Defense) and international (NATO) working groups and committees.

Jeffrey S. Palmer is the leader of the Human Health and Performance Systems Group at Lincoln Laboratory. He has oversight of multiple research programs that focus on health, human performance, objective neurocognitive analytics, and biosensing via wearable, ingestible, and implantable devices. In 2010, he helped to create the first Army War College Fellowship at Lincoln Laboratory and the MIT Security Studies Program. He has given presentations at international conferences and authored book chapters and technical articles on DNA biometrics and forensics, biomechanics, cell biology, materials science, soldier nanotechnology, biological-chemical defense, polymer science, high-energy lasers, microelectronics packaging, wearable biomedical sensing in extreme environments, and neurocognitive technologies. He has served on editorial boards for journals in biomechanics, molecular science, biomedical informatics, and biosensors. He has chaired technical conferences for the National Science Foundation, Department of Homeland Security, and the IEEE. Currently, he is the vice chair (and chair-elect) of the IEEE Engineering in Medicine and Biology Society's (EMBS) Technical Committee on Wearable Biomedical Sensors and Systems and the EMBS conference editorial board for tissue engineering and biomaterials. In addition, he has served as an advisor on senior military studies of enhancing health and performance, and led a multi-agency U.S. government effort to develop automated rapid human DNA analysis capabilities for field biometrics and forensics applications. Prior to working at Lincoln Laboratory, he worked at research laboratories at IBM and GE, and at the Physical Sciences Laboratory at New Mexico State University. He holds a bachelor's degree with a minor in mathematics from New Mexico State University, a master's degree from Rensselaer Polytechnic Institute, and a doctorate with a minor in bioengineering from MIT, all with majors in mechanical engineering.

Fundamental Brain Research

Thomas F. Quatieri, Laura J. Brattain, Todd A. Thorsen, Gregory A. Ciccarelli, Shaun R. Berry,

Jeffrey S. Palmer, Satrajit S. Ghosh, and Kwanghun Chung

The brain is a marvel of biological engineering that, if understood well, could yield unprecedented breakthroughs in treatments for neuropsychological diseases. To further develop scientists' knowledge of the brain, Lincoln Laboratory and MIT campus have teamed up in several areas of fundamental brain research. Together, we are pushing the boundaries of novel imaging and neural manipulation techniques and codifying our knowledge with neurocomputational models.

The human brain is an astoundingly complex structure that contains tens of billions of neurons connected through trillions of synapses. This circuitry is the result of our genome, development, learning, and experience, and it governs the brain's functioning. At Lincoln Laboratory, in collaboration with MIT campus, we are developing novel approaches to measuring, manipulating, and modeling the brain's structure and function. Our ultimate goal is to improve scientists' understanding of the brain and to discover advanced methods to detect, monitor, and treat neurological and neurodevelopmental conditions, such as depression, dementia, Parkinson's disease, and autism spectrum disorders. Our efforts are threefold: measuring structure through neural-level image processing, manipulating neural function by using implantable optical methods, and modeling structure and function through neurocomputational control circuits.

In our first effort, we are taking advantage of the recent advances in intact brain imaging developed by the MIT Chung Lab. These advances make it possible to collect volumetric images of brain tissue at cellular resolutions. Together with the MIT Chung Lab, we are developing robust and scalable automated neuron tracing algorithms that will provide the building blocks for developing state-of-the-art brain mapping capabilities. These capabilities will allow us to perform connectivity analysis, such as tracing specific long fibers, computing fiber density and orientations in a local region, and identifying fiber crossings.

In our second effort, we are using a powerful technique called optogenetics to monitor and control the

THOMAS F. QUATIERI, LAURA J. BRATTAIN, TODD A. THORSEN, GREGORY A. CICCARELLI, SHAUN R. BERRY, JEFFREY S. PALMER, SATRAJIT S. GHOSH, AND KWANGHUN CHUNG

functioning of populations of neurons in vivo. In collaboration with the MIT Synthetic Neurobiology Group, we are discovering methods of exciting individual neurons at arbitrary sites in the brain in a controllable, single-cell manner. This ability would open the possibility of in vivo analysis of network connectivity at the single-cell level, even in deep brain tissue that is difficult to access. We are developing an electrically controlled, micron-scale liquid lens that provides adjustable focusing and beam steering; the lens will be used for optogenetic in vivo mapping of brain activity with single-cell resolution.

In our third effort, we are applying conventional functional magnetic resonance imaging (fMRI) to modeling perception-action control of sensorimotor activities within the context of neurological disease. With our collaboration with the Gabrieli Laboratory and the Senseable Intelligence Group in the MIT McGovern Institute for Brain Research, we have developed and continue to develop scientifically grounded models through fMRI brain imaging to extract neurocomputationally

inspired biomarkers from speech and potentially other behavioral measurements. These models could improve the detection, monitoring, and phenotyping of neurological disease. We aim to transition this technology to the real world through mobile applications.

Brain Structure and Function
Anatomy

The brain is the body's most complex organ, governing motor activity, cognition, and emotion [1]. Figure 1 shows the brain's structure, illustrating the many components that make up its anatomy.

To use a computer analogy, picture the brain as the central processing unit of the body. Processing in the brain is neither exclusively digital nor analog [2]. Information coming into the brain from sensory cues (e.g., speech, audio, pressure, and pain) stimulates neural activity in the brain, with decision making being statistical rather than deterministic [3]. The brain is surrounded by an aqueous layer of cerebrospinal fluid that functions as
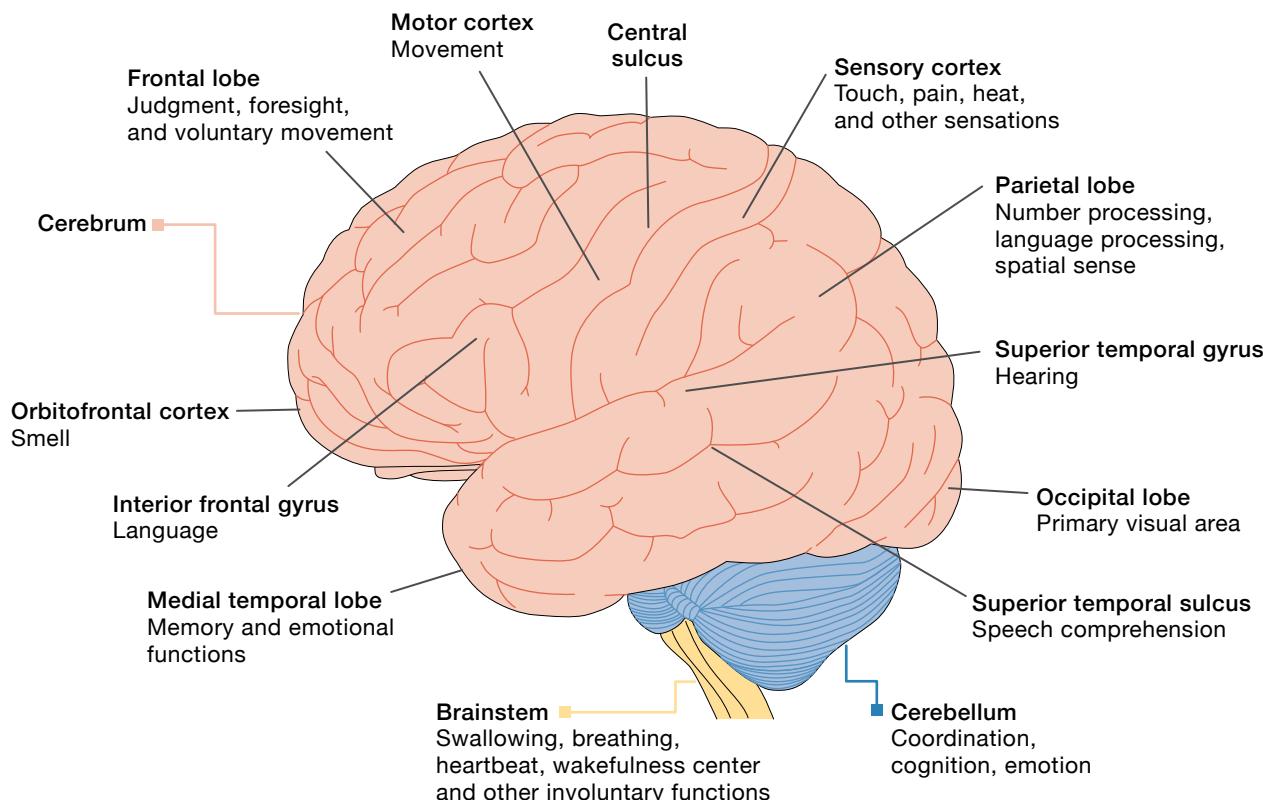
FIGURE 1. There are three principal sections of the brain: the cerebrum, cerebellum, and brainstem. Thirteen of the cerebrum's structural regions are shown along with some of their associated functions, though these regions are highly connected and can share responsibilities.

a shock absorber to mitigate trauma when the head is jarred or struck. The gross anatomy of the brain consists of three principal sections: the cerebrum, the cerebellum, and the brainstem. The largest section, occupying about two-thirds of the brain, is the cerebrum. Anatomically, it is divided into two lobes (referred to as the left brain and the right brain) containing the topical cerebral cortex layer and deeper structures, such as the hypothalamus, the olfactory bulbs, and the basal ganglia. The highly invaginated adult cerebral cortex has a surface area of about 2.5 square feet [4]. The cerebellum, occupying the rear section of the brain, is the brain's integration and coordination center, while the brainstem forms the connection with the spinal cord. The adult brain has more than 80 billion neurons along with supporting glial cells, which can be thought of as electrical insulators that prevent "short circuiting" between different neural pathways.

The left and right sides of the brain can each be further divided into four sublobes—frontal, parietal, temporal, and occipital—that play unique roles in information processing. The frontal lobe is the higher cognitive center of the brain, responsible for complex activities, such as problem solving, abstract reasoning, and moral judgment. The parietal lobe integrates sensory information such as touch, taste, and temperature and enables activities like reading and addition. The temporal lobe's central role is auditory perception, but it also plays a significant role in memory. Damage to the right and left temporal lobes has different manifestations. As examples, right-lobe damage impacts drawing and music skills, while left-lobe damage impacts memory and verbal communication. Finally, the occipital lobe is the brain's vision processing center, enabling comprehension of what the eyes are seeing.

Function and Connectivity

Understanding both anatomical and functional connectivity in the brain is important to all three areas of our brain research. Anatomical connectivity is defined by the physical connections between neurons in the brain and the strength of electrical transmission along these "wired" pathways. These regions of connectivity, called synaptic junctions, occur where the threadlike transmission regions of neurons (axons) interface with branchlike receptors (dendrites) of other neurons in close proximity. Synaptic strength, which is often used as a benchmark of anatomical connectivity, refers to the amount of current that is passed through a synaptic junction. Interestingly, the brain is continually rewiring itself, a concept often referred to as brain plasticity [5]. Anatomical mapping of the wiring between individual neural paths has shown that strong, persistent synaptic strength, measured on the order of minutes, fades because of the formation and electrical reinforcement of new synaptic junctions, which take hours to days to form [6].

In contrast to anatomical connectivity, functional connectivity can be regarded as an abstraction of the anatomical wiring of the brain. To understand functional connectivity, we look at regions of brain activity (or inactivity) rather than mapping signal transmissions along connected neurons. Techniques such as fMRI are used to map functional connectivity, imaging temporal synchronous activity fluctuations across short- and long-range regions of the brain that can be correlated with activities such as speaking, exercising, and sleeping [7]. Functional maps created from tools like fMRI are largely statistical. They show cross-correlation of activity at both gross resolutions (between different structural areas of the brain) and at millimeter scales (in specific regions like the cerebral cortex).

Large-Scale Brain Research Accelerator for Interconnected Neurons

Microscopy Background

A major goal of the U.S. government's Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative is to map the human brain at different scales with improved speed and accuracy [8]. At the macroscale, significant progress has been made in identifying long-range connectivity of axon bundles using MRI and diffusion tensor imaging [9–11]. At the microscale, electron microscopy (EM) or scanning electron microscopy (SEM) can capture fine details in extremely thin slices of the brain circuitry [12]. Although many semi-automated segmentation techniques have been developed, the processing of EM/SEM images is still limited in throughput by the machinery required to prepare the microslices and the effort required to segment and proof the content of the images [13]. To bridge the gap between macro- and microscale imaging, light-based microscopy provides mesoscale brain imaging with improved throughput [14].

THOMAS F. QUATIERI, LAURA J. BRATTAIN, TODD A. THORSEN, GREGORY A. CICCARELLI, SHAUN R. BERRY, JEFFREY S. PALMER, SATRAJIT S. GHOSH, AND KWANGHUN CHUNG

A recent significant advancement in light microscopy is the CLARITY method, which was developed for intact tissue processing [15, 16]. CLARITY (which stands for clear lipid-exchanged acrylamide-hybridized rigid imaging/immunostaining/in situ-hybridization-compatible tissue hydrogel) removes lipids that cause biological samples to be optically opaque, while preserving molecules of interest, such as proteins. This approach eliminates the need for slicing tissues into micrometer-thick sections and, for example, enables volumetric imaging of an intact mouse brain at a subcellular resolution. A successive technique, magnified analysis of proteome, or MAP [17], further improves the resolution available in intact tissues by physically expanding the tissue fourfold linearly, thus achieving super-resolution as fine as 60 nanometers (nm). This improved resolution enables light microscopy to reveal subcellular structures, such as synaptic components and intercrossing neuronal fibers in the brain.

Image Processing Approach

Despite the advances in brain imaging at different scales, dense mapping at the cellular resolution of individual neurons and neurites has been hampered by the lack of automated image processing capabilities [18]. Challenges include the large volume of data (e.g., CLARITY microscopy can generate one terabyte per hour), the density of intertwined axons and cells, and the high tracing accuracies required to avoid losing neural connections.

Existing neuron-tracing methods primarily focus on tracing single-neuron morphology [19–24], which typically includes the neural cell body (soma), axon (fiber), and branches (dendrites). A popular open-source biomedical image visualization and analysis software suite is Vaa3D [25, 26], which contains plugins that range from image filtering to neuron tracing. In our experience, we have found that Vaa3D's most recent neuron-tracing algorithms, such as Rivulet [27], are mainly designed for tracing a single neuron structure at a time, where all detected foreground pixels are considered connected. This process is different from tracing high-density neuronal fibers, such as those imaged by CLARITY/MAP, in which even closely spaced long-range axonal fibers usually are not connected.

Through Lincoln Laboratory's Large-Scale Brain Research Accelerator for Interconnected Neurons project, we are collaborating with the MIT Chung Lab to develop a new platform to enable large-scale, long-range fiber-connectivity analysis at resolutions that have not been explored. This effort is supporting the MIT Chung Lab's Integrated 3D Reconstruction of Whole Human Brains at Subcellular Resolution project, which is working with the Senseable Intelligence Group's Distributed Archives for Neurophysiology Data Integration project, as part of the BRAIN Initiative Cell Census Network, to generate and disseminate cellular-level data from humans and other species.

Initial progress is presented here on a processing pipeline for automated tracing and connectivity mapping for high-density volumetric neuronal fibers. This is the first reported work to automatically trace densely packed, immunolabeled axonal fibers in CLARITY/MAP data.

To address the unmet gap in large-scale neuron tracing, we developed the following machine learning–based, high-performance computing pipeline (Figure 2):

1. A convolutional neural network detects axon fiber voxels.
2. Morphological operations extract fiber centerlines.
3. Tracking logic connects fiber segments across low-intensity gaps and unresolved fiber crossings.

The pipeline was implemented on a CPU cluster and tested on a 250-gigabyte volume of SMI-312 densely labeled axon fibers, imaged from parts of the hippocampus and cortex of a MAP-processed mouse brain. The pipeline automatically traced 221,298 fibers across
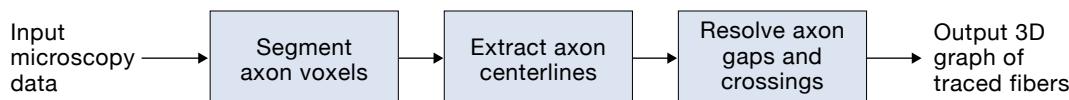
FIGURE 2. Microscopy data from mouse brain slices imaged by using MAP are taken as input into the pipeline. Voxels, or clusters of pixels in the image representing brain cells, of axon fibers are segmented, and their centerlines are determined. The system then identifies gaps and crossings between axon fibers and outputs a 3D graph of the axon fibers.

gray and white matter in 10 hours. Of the traced fibers, 104 exceeded 1 millimeter (mm), with the longest being 2.16 mm. The reported accuracy of 84 percent was based on manual evaluation of the 200 longest fibers. An example of neuron tracings is given in Figure 3.

While there is room to improve accuracy, this pipeline offers a significantly faster and more efficient method of tracing neurons than either tracing one neuron at a time or manual fiber tracing. Our pipeline can be a powerful tool for identifying and understanding the subcellular connections in the brain. The sheer number of automatically traced fibers offers a basis for high-level analysis to assess large-scale distribution statistics of long-range fiber connectivity, orientation, length, and diameter. The pipeline allows for new training data

generated from the validated fiber tracings to be fed back into the convolutional neural network model for continuous improvement of system performance. As the pipeline is scaled up to trace axon connections in larger regions of the brain, the connectivity patterns can potentially provide insight into the underlying mechanisms involved with various brain disorders.

Optogenetics to Control Neural Circuitry Manipulation Background

In the last decade, optogenetics has emerged as a powerful tool to monitor and control the functioning of neuron populations in vivo [28]. Optogenetics involves selectively photoexciting neurons that are genetically modified to express photosensitive membrane proteins (opsins) [29].
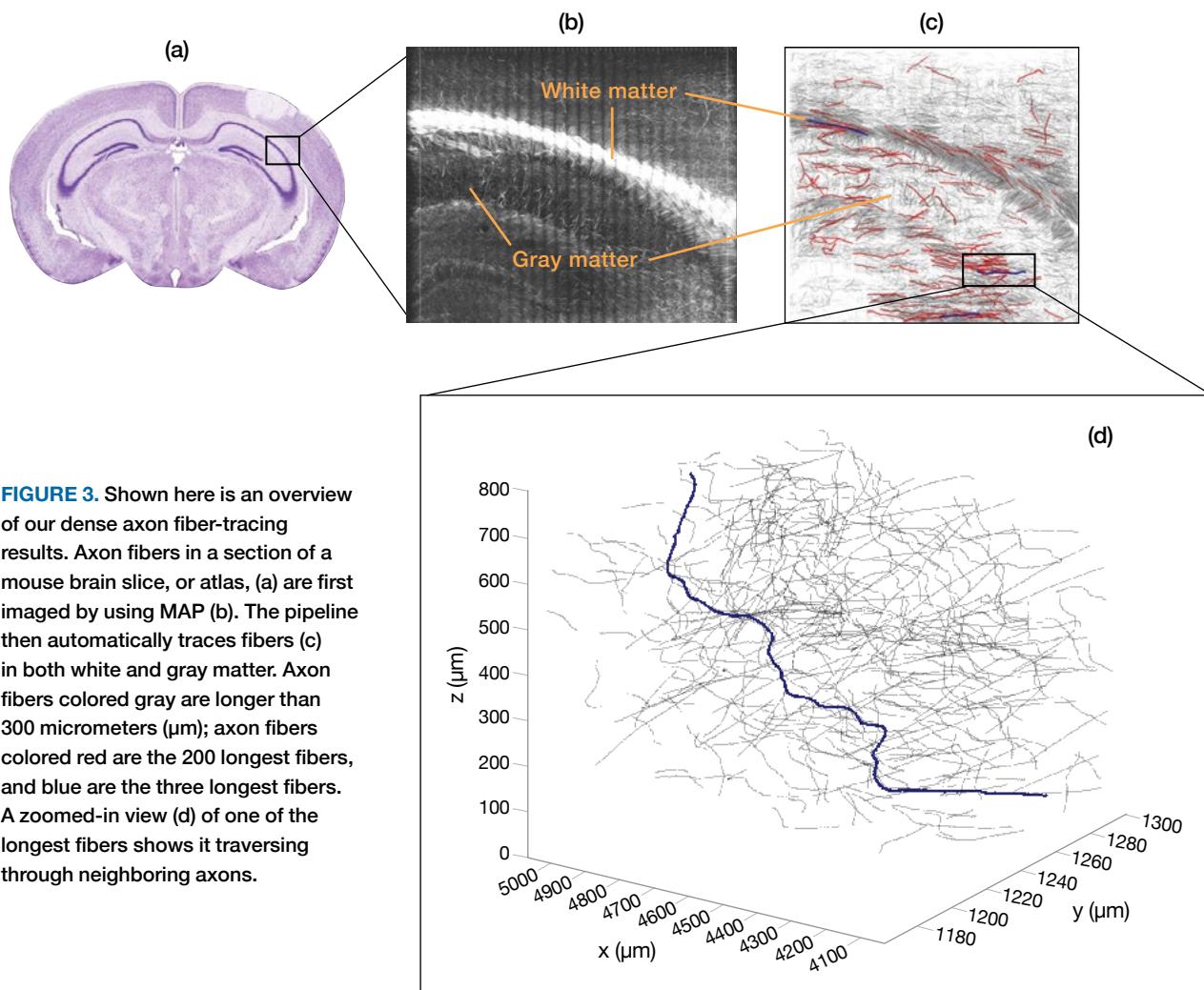
FIGURE 3. Shown here is an overview of our dense axon fiber-tracing results. Axon fibers in a section of a mouse brain slice, or atlas, (a) are first imaged by using MAP (b). The pipeline then automatically traces fibers (c) in both white and gray matter. Axon fibers colored gray are longer than 300 micrometers (μm); axon fibers colored red are the 200 longest fibers, and blue are the three longest fibers. A zoomed-in view (d) of one of the longest fibers shows it traversing through neighboring axons.

THOMAS F. QUATIERI, LAURA J. BRATTAIN, TODD A. THORSEN, GREGORY A. CICCARELLI, SHAUN R. BERRY, JEFFREY S. PALMER,

SATRAJIT S. GHOSH, AND KWANGHUN CHUNG

Upon excitation, these opsins transport ions into or out of neurons to control their electrical activity. So far, most optogenetics studies have involved photoexcitation with limited spatial resolution, relying on orthogonal genetically modified channelrhodopsins. These opsins respond to different colors of light that are selectively expressed in different classes of neurons [30]. Two-photon control is possible but only within the shallow depths (i.e., less than 1 mm) afforded by two-photon penetration into the brain [31–33]. Thus, it is highly desirable to have a method of exciting individual neurons at arbitrary sites in the brain in a controllable and single-cell manner because this method would open the possibility of in vivo analysis of network connectivity at the single-cell level, even in deep brain tissue that is difficult to access.

Our ultimate goal is to develop an implantable optical probe that has active focusing and steering optics placed at the end of the probe to enable light delivery from an external laser to individual neurons. We have made progress in the first phase of our work, namely engineering micron-scale liquid lenses with active electronics to enable both focusing and steering. These lenses are designed in such a way that, in later phases of the program, they can be integrated with waveguides operating in the red and near-infrared spectra, building from previous work we have done on microfabricated waveguide-mediated optogenetic control [34]. This integration will enable two-photon optogenetic excitation of individual cells at arbitrary locations in living mammalian brains.

Fabrication Approach

We employ optofluidics, whereby microlenses with diameters as small as 50 micrometers (μm) are formed by shaping the interface between two immiscible liquids with different refractive indexes. Electrowetting [35] is used to control the shape of the liquid-liquid interface, thus providing variable focus and beam-steering functionalities. With electrowetting, the surface energy of the solid substrate is modified by the application of a voltage, changing it from hydrophobic to hydrophilic in prescribed regions via suitably designed electrodes covered by a hydrophobic film. Previously, Lincoln Laboratory developed liquid, plano-convex, adjustable-focal-length microlenses with electrode diameters as small as 200 μm [36]. In our current work, a novel design of the
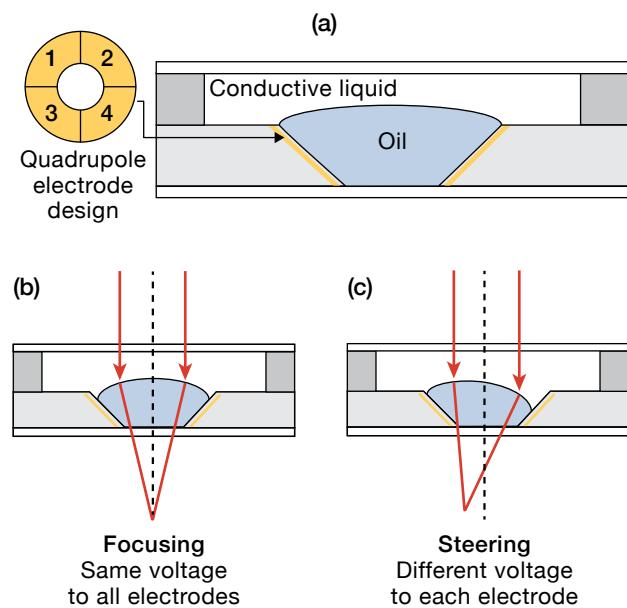
FIGURE 4. The liquid microlens design combines both active focusing and beam steering by controlling the interface formed between two immiscible liquids (a). The liquid interface is contained within a 45° conical taper that has a series of patterned metal electrodes etched along the sidewall. Through electrowetting, the interface curvature changes as a function of applied voltage to the electrodes. Variable focusing (b) along the optical path will occur when the same potential is applied to all the electrodes. Beam steering (c) will occur when different potentials are applied to each electrode.

substrate and electrodes has allowed for beam steering and a significant reduction in the microlens diameter.

A target area for excitation that contains 100,000 neuron cell bodies would require a microlens with an adjustable focal length of 0.1–1 mm and simultaneous steering over ± 5 degrees, and capable of delivering light to a spot size of ~10 μm. To minimize the overall size of the optic, focusing and steering adjustments are made via a single optical element. Both functionalities are achieved by embedding the liquid interface in a conical taper, which has interdigitated electrodes patterned along the sidewall and is etched into a fused silica substrate, as shown in Figure 4. Initial designs used either four independent electrodes or a single electrode around the conical taper. For electrowetting to work, one of the liquids needs to be conductive, typically water, and the other liquid needs to be insulating, typically a nonpolar solvent like oil. The positioning of the oil and water forms the microlens and is controlled by a patterned hydrophobic film over the

(a)   (b)
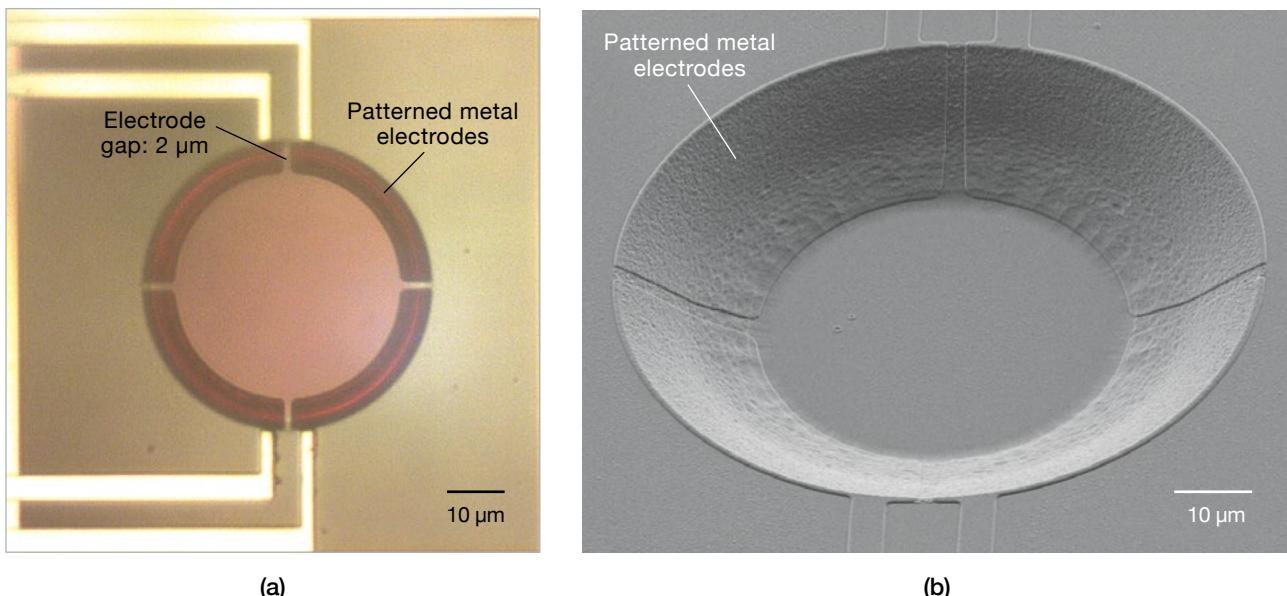
FIGURE 5. The top-down microscope image of the microlens cavity (a) shows the electrode metal deposition and patterning. The scanning electron microscopy image (b) was taken after the microlens was fabricated (the CYTOP film had not yet been applied).

electrodes and a surrounding hydrophilic surface. Oil resides inside the taper over the patterned hydrophobic film. When a voltage is applied, the surface energy changes from hydrophobic to more hydrophilic, allowing water to wet over the electrode. This transition changes the radius of curvature of the liquid interface and thus the focal length of the microlens.

The fabricated microlenses have a 12-µm-deep, conical-tapered cavity with a 45-degree sloping sidewall that was etched into a 750 µm fused silica wafer by using a gray-scale lithography technique [37]. A special optical mask designed with subresolution-sized features created a tapered profile in photoresist. The tapered pattern was then transferred into the fused silica through an optimized reaction-ion etching process. Next, 250 nm of aluminum were deposited and etched to form the electrodes. Some microlenses had a quadrupole arrangement. Others had just a single circular electrode, which allows only for adjusting focus in the microlens. After the electrodes were patterned, 500 nm of plasma-enhanced chemical vapor deposition (PECVD) oxide were deposited, forming the dielectric layer. The final fabrication step was to create the hydrophobic and hydrophilic regions. We used 80-nm-thick CYTOP (Asahi Glass) as the hydrophobic film. The film was patterned and etched in oxygen plasma to form hydrophobic regions inside the

taper over the electrodes, leaving hydrophilic regions elsewhere. Figure 5 shows images from the microfabrication process.

To conveniently evaluate the microlens performance, we designed a 10 × 10 mm test chip containing 44 individual microlenses. A completely self-contained microlens package was developed to enable optical characterization. To form the microlenses, the chip was lowered into a beaker containing the conducting liquid with a thin film of oil on the top. Through self-assembly, oil remained on the hydrophobic regions surrounded by the conducting liquid as the chip was lowered into the beaker. While the chip was submerged, a fused silica cap was lowered over it, encapsulating the microlenses. The chip was removed from the beaker, dried, and sealed with epoxy. The sealed chip was wire bonded in a 40-pin dual in-line package, which had a through-hole drilled in the socket region. Finally, the packaged microlens chip was inserted into a zero-insertion-force board-mounted connector to allow for optical characterization.

Beamforming for Manipulation

To measure the focal length of the liquid microlenses as a function of applied voltage, we developed a novel approach based on beam magnification. The focal length is determined by the change in beam magnification that

THOMAS F. QUATIERI, LAURA J. BRATTAIN, TODD A. THORSEN, GREGORY A. CICCARELLI, SHAUN R. BERRY, JEFFREY S. PALMER, SATRAJIT S. GHOSH, AND KWANGHUN CHUNG
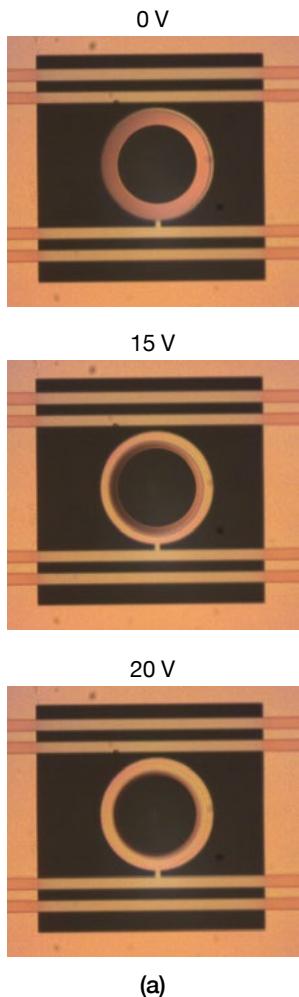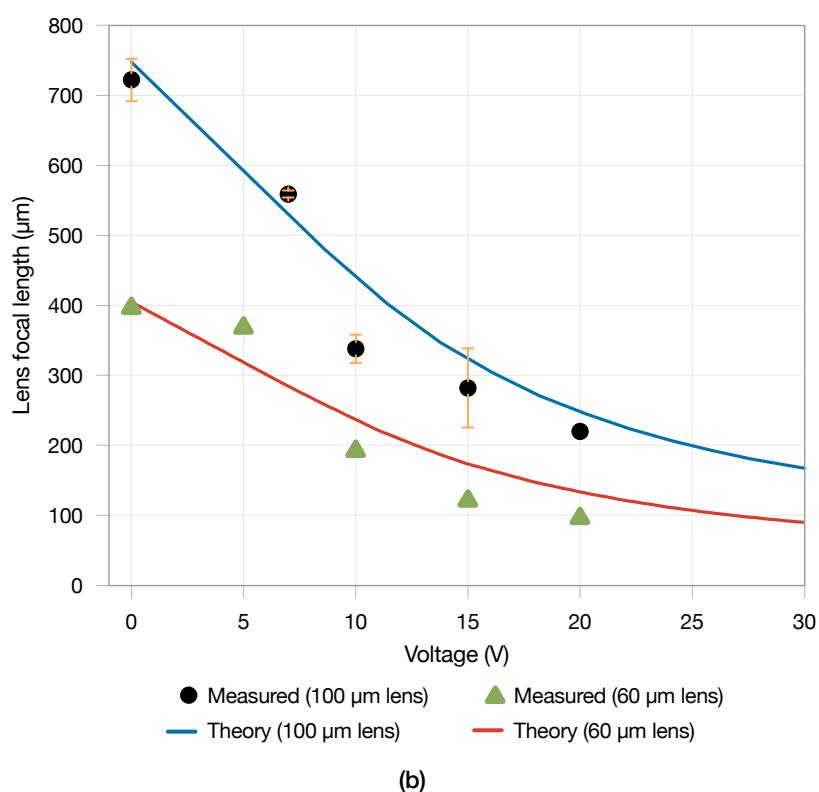
FIGURE 6. In (a), top-down microscope images show the shape of the microlens, which has a diameter of 100 μm, at different applied voltages. The graph in (b) plots the focal length versus voltage for fully packaged, planar, single-electrode, liquid microlenses. The microlenses were sealed with deionized water and DC-704 silicone oil.

the microlens imparts on a predetermined beam magnification optical system. The optical diagnostic system focuses a diagnostic beam through the test microlens with an ~16 μm spot size and outputs a beam with a size compatible with a beam-profiling camera. In principle, if the microlens is aligned precisely in the optical diagnostic system, then its focal length can be determined directly from the measured beam size. However, determining the microlens' focal length is difficult in practice. Consequently, the microlens was also translated along the optical axis of the diagnostic system, similar to a z-scan, with beam size measurements recorded at multiple positions. These additional measurements provide sufficient data to relax the alignment tolerance of the microlens, while improving the sensitivity of the focal length measurement. Initial characterization was done by using a helium–neon laser (633 nm wavelength). The microlens focal length was determined by fitting the measured beam size as a function of translated distance to an exact analytical expression.

Figure 6 shows the focal-length-versus-voltage results from characterizing two fully packaged, planar, single-electrode, liquid microlenses of different sizes. As the voltage is increased, each lens becomes more convex as water wets the electrodes and forces the oil into the center of the aperture, decreasing the effective focal length. Through active control of the focal length, the lenses can be ultimately steered to focus light on individual neurons at different depths in the brain.

Neurocomputational Modeling
Modeling Background
Neurocomputational modeling is the science of constructing and using mathematical models of neuro-biophysical processes to characterize brain function [38]. The current standard of care in neuropsychiatry

has not caught up with the wealth of neuroscientific data that new imaging and manipulation techniques have made available. There is a pressing need for a unified, actionable framework for knowledge that is mined from state-of-the-art imaging techniques and ubiquitous data collection technologies, such as smartphones. The challenges are numerous: data are sparse, multimodal, and noisy, and the underlying sensorimotor and cognitive processes at work are seemingly impenetrable in their complexity. Appropriate modeling strategies can address these challenges and have demonstrated potential on real-world data.

Modeling can be divided into two stages: model construction and model use. In model construction, a mathematical summary of relationships between quantitative data is created. Quantitative data are aggregated from subjects under study (e.g., people with Alzheimer's disease, dementia, traumatic brain injury, autism, depression, or Parkinson's disease) and can include modalities such as voice, structural and functional MRI, accelerometer gait data, and known medications. These data are fused with prior knowledge of neuroanatomical function based on previous human and animal studies. Often, a complete prior model does not exist for the data and disorder under study, so model construction also involves a knowledge-discovery component. Modelers must choose between competing possible hypotheses by using their prior knowledge or tools such as Bayesian information criterion and out-of-sample prediction.

Model use is the leveraging of an individual's fitted model to understand and provide treatment for that individual. Models can be used to predict brain responses (e.g., dynamic causal modeling), effectiveness when sleep deprived (e.g., the Sleep, Activity, Fatigue, and Task Effectiveness [SAFTE] model), treatment response (e.g., how neural circuits can be advantageously altered in Parkinson's disease through deep brain stimulation), and neurotraumatic damage. Models also are a compact summary of the subject's data. Consequently, differences between individual models can be used for disease severity assessment and trajectory prediction. The mathematical constants that parameterize a model, whether those are time constants of neural processes or strength and prevalence of functional connectivity, are all potential features for classical machine learning algorithms. Machine learning algorithms trained on these parameters can then predict responses potentially better than or complementary to those trained on the raw data.

Neurocomputational modeling seeks a mathematical, mechanistic explanation for behavior through the use of observations and hypothesized biophysical and neurological mechanisms. An example of this framework in action is a unifying paradigm of motor control that can be illustrated with a specific example of speech production and the neurological disorders of depression and Parkinson's disease. Our modeling allows a complementary approach to traditional feature engineering schemes, with model features reflecting internal or latent model parameters. This approach is a step toward a brain basis for biomarkers of a disease and thus toward increased clinical acceptance of behavior-based automatic assessment systems [38].

Imaging as a Modeling Basis

Brain imaging can reveal neurophysiological structures that are relevant for speech, and these structures act as a starting point for neurocomputational models of speech production. Speech production results from the coordinated activity of a distributed set of neuroanatomical regions, as has been determined through various forms of brain imaging. Figure 7 shows examples of fMRI images of some primary regions of the speech network. These regions, determined through speech-production-task studies [39], include cortical and subcortical components as well as the cerebellum and brainstem, and involve feedforward and feedback mechanisms in prosodic, articulatory, and linguistic components of speech production.

Specifically, the posterior superior temporal gyrus and Heschl's gyrus are part of the auditory cortex and are crucial for self-monitoring in speech production feedback. The subcortical basal ganglia (composed in part by the striatum) and the midline cortical areas of the supplementary motor area and presupplementary motor area are responsible for initiation and plan selection and speech sequencing. The motor cortex, specifically the ventral motor cortex, and parts of the inferior frontal gyrus (Broca's area) control the articulators for the actual production of speech. The cerebellum further assists in the fine timing control of articulators needed for well-formed sounds. Prosodic modulation relies on the core speech network but also recruits part

THOMAS F. QUATIERI, LAURA J. BRATTAIN, TODD A. THORSEN, GREGORY A. CICCARELLI, SHAUN R. BERRY, JEFFREY S. PALMER, SATRAJIT S. GHOSH, AND KWANGHUN CHUNG
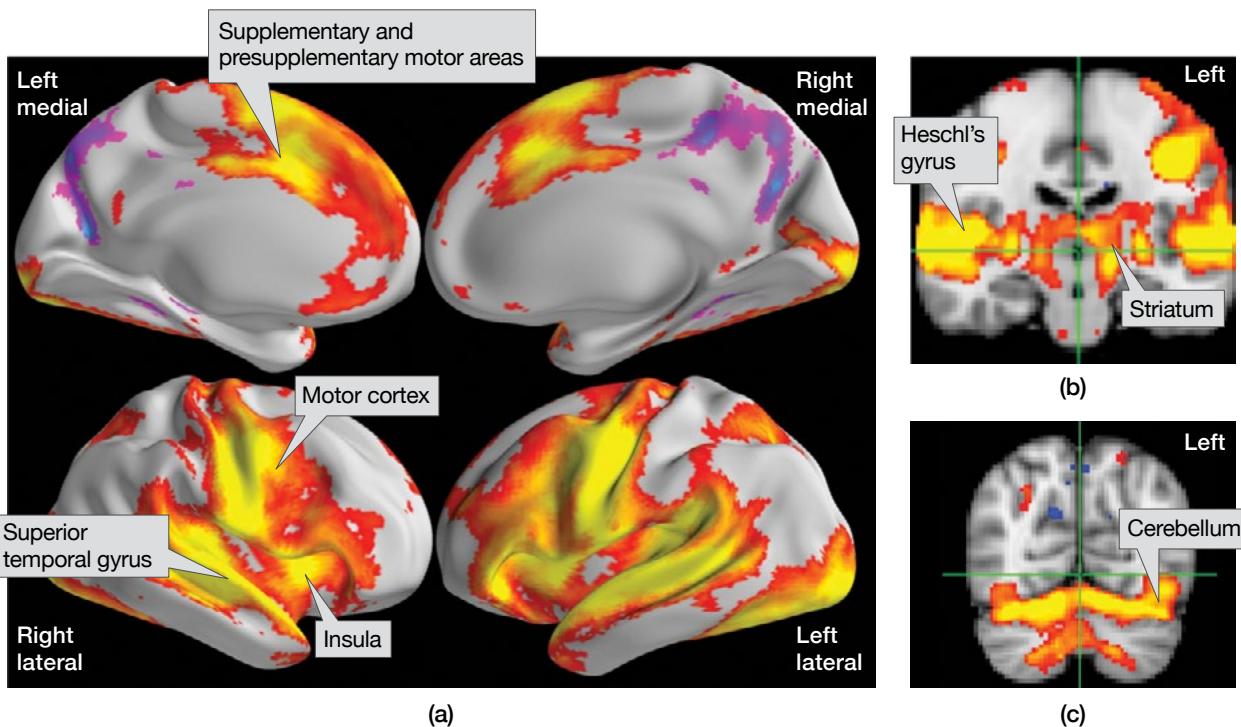
FIGURE 7. These functional magnetic resonance images show speech production–related regions of interest in the brain. The four images in (a) show activity in the cortex. The medial views of the brain at top (a) show the supplementary motor area and presupplementary motor area used for speech planning and sequencing. At bottom (a), a lateral view of the brain shows bilateral activation of the motor cortex used for controlling the articulators of speech production and shows the superior temporal gyrus, part of the auditory cortex, used for self-monitoring speech production. The coronal view (b) of the brain shows Heschl's gyrus, which is involved in auditory processing, and the striatum, which with the insula is hypothesized to connect limbic processing to the speech production system. The coronal view (c) of the posterior brain shows bilateral activation in the cerebellum. The cerebellum assists in precisely timing motor commands in speech production [42].

of the limbic system (e.g., the amygdala and the insula) in emotional speech. F.H. Guenther [39], C.J. Price [40], and S. Pichon et al. [41] provide further details on the neuroanatomy of speech production.

Some of these speech regions overlap or are connected with neuroanatomical regions, such as the amygdala or motor regions, that are associated with our example conditions of depression and Parkinson's disease, respectively. Details of these links are out of the scope of this article (please see the article "Noninvasive Biomarkers of Neurobehavioral Performance" on page 28 in this *Journal*), but essentially these connections may provide an opportunity for nonspeech processes to influence speech production. The hypothesized modulation of the speech network by nonspeech processes and deficits that can occur in modules of the speech production network provide the guiding motivation for why

speech can be a biomarker of neuropsychological disorders in general and of depression and Parkinson's disease in particular.

Although not a focus of this article, Lincoln Laboratory's Human Health and Performance Systems Group is developing mechanism-based modeling approaches to enhance neurotrauma assessments. Imaging is also useful as a structural input to establish neurophysiological-based models, as well as for bioelectric and biomechanical constitutive models of the various hard and soft tissues and fluids that compose brain structures. Energetic impulses can deliver direct traumatic insults to the brain and can be converted to other forms of damaging input (e.g., electromagnetic energy to thermomechanical force) [43]. The time- and temperature-dependent nature of the damaging inputs and the pathophysiological response within the brain

require these neurotrauma models to be based not solely on linear properties [44]. Nonlinear constitutive models (e.g., viscoelasticity, hyperelasticity, and poroelasticity) have been used for high-fidelity damage assessments from photothermal, thermomechanical, and overpressure loads [45]. Our models can input multiple two-dimensional MRI scans to create a three-dimensional neurostructural model with varying nonlinear material properties for ventricles, glia, white matter, gray matter, eyes, venous sinus, cerebrospinal fluid, air sinus, muscle, skin, fat, and tribecular bone (e.g., cranium) [46].

Perception-Action Framework

Brain imaging representations of the type shown in Figure 7 have motivated neurocomputational models of motor control [39]. A unifying paradigm of motor control is the perception-action framework diagrammed in Figure 8 [47]. We describe a neurophysiological system through several broad components and relationships. The first component is the biophysical plant, or motor system, that is controlled by the controller module, the second component. The third component is a sensing module responsible for perception, and the fourth component is the goal, or plan, responsible for the desired action. While precise delineation of roles is debatable, we will determine for this article that the sensory system detects the state of the plant, and that this current state and the goal—the desired state of the plant as instructed by higher-level cognitive control—are both inputs to the controller. We note that the goal of the system may have a different parameterization than the plant itself has. For example, in speech production the goal may be an auditory target, but the plant configuration that corresponds to such a goal could be specified as a function of articulator positions. The controller then implements corrective action by taking the error between the desired and perceived states and sends a control signal to the biophysical plant. The controller closes the loop between perception and action.

Importantly, the perception-action framework has an "observable biomarker" analog pictured in Figure 9. Although we cannot generally see the workings of the system under study, we can observe various biomarkers that can be used to tune a model of the plant to an approximate match of the real state. Therefore, we use observable biomarkers to provide system identification on what, without a modeling framework, would be a black box.
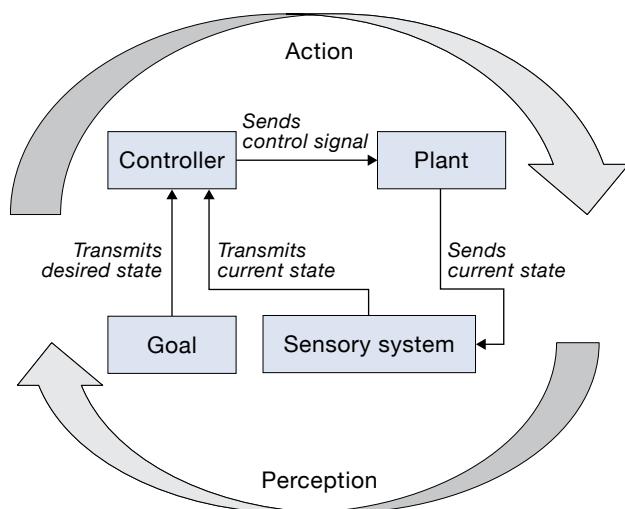
FIGURE 8. A neurocomputational modeling framework is based on the perception-action loop, which allows the brain to seamlessly integrate goals, sensory perceptions, and new motor commands to control the plant (i.e., the body) [47].

Applying Neurocomputational Modeling to Assessing Neuropsychological Disorders

In research by Ciccarelli et al., neurocomputational modeling was implemented with respect to speech motor control and the neurological conditions of depression and Parkinson's disease [42, 48, 49]. Speech is one of the most complex and demanding motor tasks and draws upon a significant amount of neural "real estate" in the brain. Consequently, speech production biomarkers may be sensitive to neurological disorders. A computational model of speech production that follows the described perception-action paradigm is the Directions into Velocities of Articulators (DIVA) model developed by Guenther et al. [39, 50]. This model describes the workings of the vocal tract resonances that are responsible for creating the acoustic cues that differentiate vowel sounds.

Ciccarelli et al. adopted this particular perception-action framework to describe the vocal folds that are responsible for creating fundamental frequency, which is perceived as pitch [49]. This adoption required leveraging a biophysical model of the vocal folds for the plant, with the vocal source model characterized by two latent muscle parameters, the cricothyroid (CT) and thyroarytenoid (TA) muscles of the larynx. The coordination of the two estimated muscle intensities over time was then used to characterize a disease, with coordination

THOMAS F. QUATIERI, LAURA J. BRATTAIN, TODD A. THORSEN, GREGORY A. CICCARELLI, SHAUN R. BERRY, JEFFREY S. PALMER, SATRAJIT S. GHOSH, AND KWANGHUN CHUNG
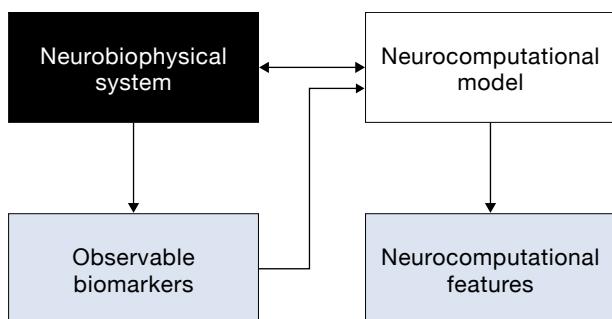
FIGURE 9. The black box neurobiophysical system can be investigated with a white box neurocomputational model by estimating model parameters and structure from observable biomarkers. The model can then be used to derive neurocomputational features such as muscle intensities and error signals [49].

represented by multiscale correlation of the CT and TA time series [49, 53]. Ciccarelli et al. embedded this plant representation within the perception-action framework to perform system identification (i.e., the unknown CT and TA muscle intensities) on individuals with depression and Parkinson's disease [49]. The underlying features estimated in this way were then used as biomarkers to assess the presence or severity of the underlying disorder.

Results and Discussion

Ciccarelli et al. [49] applied these features to the depression database of the Audio/Visual Emotion Challenge [51], a competition that tasks participants with estimating the level of subjects' depression by using audio and visual data. Compared to traditional, non-biophysically motivated features, the neurocomputational biophysical features provided nearly twice the explanatory power for depression. Ciccarelli et al. also applied this technique to predicting the severity of Parkinson's disease by using the mPower database [52], a collection of health-condition data and symptom-progression data from Parkinson's disease patients. They found that although the model-based features alone underperformed more traditional features, fusion of the two predictors performed better than either alone.

A similar neurocomputational modeling–based approach was also introduced by Ciccarelli [42] and Williamson et al. [48] for vocal tract movement. This approach illustrated the coordination over time of the

vocal tract muscle intensities (responsible for coordinating vocal tract resonances) used to characterize Parkinson's disease. Again, fusing the model-based features with traditional features provided more benefit than either feature set alone in prediction of severity of Parkinson's disease [48].

In summary, we advocate a neurobiophysical model embedded within a perception-action framework. A mathematical, mechanistic description of a person can provide new biomarkers through system identification and permit experimental manipulations via computer simulation to potentially provide further insight into neuropsychological disorders, such as depression, Parkinson's disease, traumatic brain injury, Alzheimer's disease, and dementia. Such modeling may lead to more accurate and early detection of a neurological disorder, phenotyping of the disorder, monitoring of treatment effects, and predictions of new therapy outcomes.

Synergy Moving Forward

We have described three areas of brain-related research at Lincoln Laboratory in collaboration with MIT campus: measuring structure through neural-level image processing, manipulating neural function by using implantable optical methods, and modeling structure and function through neurocomputational control circuits.

Though these three efforts may appear disjointed, they are potentially strongly synergistic. Accurate neurocomputational models will rely on more highly resolved connectivity maps obtained through refined neural tracing, while increased understanding of the function of feedforward and feedback model pathways will rely on the manipulation of neural circuits through more advanced optogenetic-based techniques. We believe this synergy will lead to an improved neural mechanistic explanation for behavior and, in particular, improved phenotyping across and within neurological disorders. ■

References

1. L. Pessoa, "The Cognitive-Emotional Brain: From Interactions to Integration," *Neuropsychologia*, vol. 48, no. 12, 2010, pp. 3416–3429.

2. Y. Mochizuki and S. Shinomoto, "Analog and Digital Codes in the Brain," *Physical Review E*, vol. 89, no. 2, 2014, article no. 022705.

3. J.I. Sanders, B. Hangya, and A. Kepecs, "Signatures of a Statistical Computation in the Human Sense of Confidence," *Neuron*, vol. 90, no. 3, 2016, pp. 499–506.

4. A. Peters and E.G. Jones (eds.), *Cerebral Cortex. Vol 1, Cellular Components of the Cerebral Cortex*. New York: Plenum Press, 1984.

5. H. Johansen-Berg, "Structural Plasticity: Rewiring the Brain," *Current Biology*, vol. 17, no. 4, 2007, pp. R141–R144.

6. L.C. Katz and C.J. Shatz, "Synaptic Activity and the Construction of Cortical Circuits," *Science*, vol. 274, no. 5290, 1996, pp. 1133–1138.

7. N. Kriegeskorte, R. Goebel, and P. Bandettini, "Information-Based Functional Brain Mapping," paper in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 10, 2006, pp. 3863–3868.

8. "Brain Research through Advancing Innovative Neurotechnologies (BRAIN) —National Institutes of Health (NIH)," available at https://www.braininitiative.nih.gov/.

9. O. Sporns, "The Human Connectome: a Complex Network," *Annals of the New York Academy of Science*, vol. 1224, no. 1, 2011, pp. 109–125.

10. A. Zalesky, A. Fornito, I.H. Harding, L. Cocchi, M. Yücel, C. Pantelis, et al., "Whole-Brain Anatomical Networks: Does the Choice of Nodes Matter?" *NeuroImage*, vol. 50, no. 3, 2010, pp. 970–983.

11. J.C. Shillcock, M. Hawrylycz, S. Hill, and H. Peng, "Reconstructing the Brain: From Image Stacks to Neuron Synthesis," *Brain Informatics*, vol. 3, no. 4, 2016. pp. 205–209.

12. M. Helmstaedter, K.L. Briggman, S.C. Turaga, V. Jain, H.S. Seung, and W. Denk, "Connectomic Reconstruction of the Inner Plexiform Layer in the Mouse Retina," *Nature*, vol. 500, no. 7461, 2013, pp. 168–174.

13. I. Arganda-Carreras, S.C. Turaga, D.R. Berger, D. Cireşan, A. Giusti, L.M. Gambardella, et al., "Crowdsourcing the Creation of Image Segmentation Algorithms for Connectomics," *Frontiers in Neuroanatomy*, vol. 9, 2015, p. 142.

14. M.B. Ahrens, M.B. Orger, D.N. Robson, J.M. Li, and P.J. Keller, "Whole-Brain Functional Imaging at Cellular Resolution Using Light-Sheet Microscopy," *Nature Methods*, vol. 10, no. 5, 2013, pp. 413–420.

15. K. Chung and K. Deisseroth, "CLARITY for Mapping the Nervous System," *Nature Methods*, vol. 10, no. 6, 2013, pp. 508–513.

16. K. Chung, J. Wallace, S.-Y. Kim, S. Kalyanasundaram, A.S. Andalman, T.J. Davidson, et al., "Structural and Molecular Interrogation of Intact Biological Systems," *Nature*, vol. 497, no. 7449, 2013, pp. 332–337.

17. T. Ku, J. Swaney, J.-Y. Park, A. Albanese, E. Murray, J.H. Cho, et al., "Multiplexed and Scalable Super-Resolution Imaging of Three-Dimensional Protein Localization in Size-Adjustable Tissues," *Nature Biotechnology*, vol. 34, no. 9, 2016, pp. 973–981.

18. J.W. Lichtman, H. Pfister, and N. Shavit, "The Big Data Challenges of Connectomics," *Nature Neuroscience*, vol. 17, no. 11, 2014, pp. 1448–1454.

19. Y. Wang, A. Narayanaswamy, C.-L. Tsai, and B. Roysam, "A Broadly Applicable 3-D Neuron Tracing Method Based on Open-Curve Snake," *Neuroinformatics*, vol. 9, no. 2–3, 2011, pp. 193–217.

20. E. Türetken, G. González, C. Blum, and P. Fua, "Automated Reconstruction of Dendritic and Axonal Trees by Global Optimization with Geometric Priors," *Neuroinformatics*, vol. 9, no. 2–3, 2011, pp. 279–302.

21. P. Chothani, V. Mehta, and A. Stepanyants, "Automated Tracing of Neurites from Light Microscopy Stacks of Images," *Neuroinformatics*, vol. 9, no. 2–3, 2011, pp. 263–278.

22. H. Peng, F. Long, and G. Myers, "Automatic 3D Neuron Tracing Using All-Path Pruning," *Bioinformatics*, vol. 27, no. 13, 2011, pp. i239–i247.

23. E. Bas and D. Erdogmus, "Principal Curves as Skeletons of Tubular Objects," *Neuroinformatics*, vol. 9, no. 2–3, 2011, pp. 181–191.

24. M.H. Longair, D.A. Baker, and J.D. Armstrong, "Simple Neurite Tracer: Open Source Software for Reconstruction, Visualization and Analysis of Neuronal Processes," *Bioinformatics*, vol. 27, no. 17, 2011, pp. 2453–2454.

25. H. Peng, Z. Ruan, F. Long, J.H. Simpson, and E. W. Myers, "V3D Enables Real-Time 3D Visualization and Quantitative Analysis of Large-Scale Biological Image Data Sets," *Nature Biotechnology*, vol. 28, no. 4, 2010, pp. 348–353.

26. H. Peng, A. Bria, Z. Zhou, G. Iannello, and F. Long, "Extensible Visualization and Analysis for Multidimensional Images Using Vaa3D," *Nature Protocols*, vol. 9, no. 1, 2014, pp. 193–208.

27. S. Liu, D. Zhang, S. Liu, D. Feng, H. Peng, and W. Cai, "Rivulet: 3D Neuron Morphology Tracing with Iterative Back-Tracking," *Neuroinformatics*, vol. 14, no. 4, 2016, pp. 387–401.

28. K. Deisseroth, G. Feng, A. Majewska, G. Miesenböck, A. Ting, and M. Schnitzer, "Next-Generation Optical Technologies for Illuminating Genetically Targeted Brain Circuits," *Journal of Neuroscience*, vol. 26, no. 41, 2006, p. 10380.

29. E. Boyden, F. Zhang, E. Bamberg, G. Nagel, and K. Deisseroth, "Millisecond-Timescale, Genetically Targeted Optical Control of Neural Activity," *Nature Neuroscience*, vol. 8, no. 9, 2005, pp. 1263–1268.

30. N.C. Klapoetke, Y. Murata, S.S. Kim, S.R. Pulver, A. Birdsey-Benson, Y.K. Cho, et al., "Independent Optical Excitation of Distinct Neural Populations," *Nature Methods* vol. 11, 2014, pp. 338–346.

31. J.P. Rickgauer, and D.W. Tank, "Two-Photon Excitation of Channelrhodopsin-2 at Saturation," paper in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 35, 2009, pp. 15025–15030.

32. B.K. Andrasfalvy, B.V. Zemelman, J. Tang, and A. Vaziri, "Two-Photon Single-Cell Optogenetic Control of Neuronal Activity by Sculpted Light," paper in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 26, 2010, pp. 11981–11986.

THOMAS F. QUATIERI, LAURA J. BRATTAIN, TODD A. THORSEN, GREGORY A. CICCARELLI, SHAUN R. BERRY, JEFFREY S. PALMER, SATRAJIT S. GHOSH, AND KWANGHUN CHUNG

33. E. Papagiakoumou, F. Anselmi, A. Bègue, V. de Sars, J. Glücksta, E.Y. Isacoff, and V. Emiliani, "Scanless Two-Photon Excitation of Channelrhodopsin-2," *Nature Methods*, vol. 7, no. 10, 2010, pp. 848–854.

34. A.N. Zorzos, J. Scholvin, E.S. Boyden, and C.G. Fonstad, "Three-Dimensional Multiwaveguide Probe Array for Light Delivery to Distributed Brain Circuits," *Optics Letters*, vol. 37, no. 23, 2012, pp. 4841–4843.

35. F. Mugele and J.-C. Baret, "Electrowetting: From Basics to Applications," *Journal of Physics: Condensed Matter*, vol. 17, no. 28, 2005, pp. R705–R774.

36. S. Berry, J. Stewart, T. Thorsen, and I. Guha, "Development of Adaptive Liquid Microlenses and Microlens Arrays," paper in MOEMS and Miniaturized Systems XII, *Proceedings of SPIE*, vol. 8616, 2013, pp. 861610–8616101.

37. C.M. Waits, R. Ghodssi, M.H. Ervin, and M. Dubey, "MEMS-Based Gray-Scale Lithography," paper in *2001 International Semiconductor Device Research Symposium Proceedings*, 2001, pp. 182–185.

38. T.V. Wiecki, J. Poland, and M.J. Frank, "Model-Based Cognitive Neuroscience Approaches to Computational Psychiatry Clustering and Classification," *Clinical Psychological Science*, vol. 3, no. 3, 2015, pp. 378–399.

39. F.H. Guenther, *Neural Control of Speech*. Cambridge, Mass.: MIT Press, 2016.

40. C.J. Price, "A Review and Synthesis of the First 20 years of PET and fMRI Studies of Heard Speech, Spoken Language and Reading," *NeuroImage*, vol. 62, no. 2, 2012, pp. 816–847.

41. S. Pichon and C.A. Kell, "Affective and Sensorimotor Components of Emotional Prosody Generation," *Journal of Neuroscience*, vol. 33, no. 4, 2013, pp. 1640–1650.

42. G. Ciccarelli, "Characterization of Phoneme Rate as a Vocal Biomarker of Depression," Dissertation, MIT, 2017.

43. N.M. Yitzhak, R. Ruppin, and R. Hareuveny, "Numerical Simulation of Pressure Waves in the Cochlea Induced by a Microwave Pulse," *Bioelectromagnetics*, vol. 35, no. 7, 2014, pp. 491–496.

44. T.P. Prevost, A. Balakrishnan, S. Suresh, and S. Socrate, "Biomechanics of Brain Tissue," *Acta Biomaterialia*, vol. 7, no. 1, 2011, pp. 83–95.

45. D.F. Moore, A. Jérusalem, M. Nyein, L. Noels, M.S. Jaffee, and R.A. Radovitzky, "Computational Biology—Modeling of Primary Blast Effects on the Central Nervous System," *NeuroImage*, vol. 47, 2009, pp. T10–T20.

46. A. Jeana, M.K. Nyeina, J.Q. Zheng, D.F. Moore, J.D. Joannopoulos, and R. Radovitzky, "An Animal-to-Human Scaling Law for Blast-Induced Traumatic Brain Injury Risk Assessment," paper in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 43, 2014, pp. 15310–15315.

47. V. Cutsuridis, A. Hussain, and J.G. Taylor, *Perception-Action Cycle: Models, Architectures, and Hardware*. New York: Springer, 2011.

48. J.R. Williamson, T.F. Quatieri, B.S. Helfer, J. Perricone, S.S. Ghosh, G. Ciccarelli, and D.D. Mehta, "Segment-Dependent Dynamics in Predicting Parkinson's Disease," paper presented at INTERSPEECH 2015: 16th Annual Conference of the International Speech Communication Association, 2015.

49. G.A. Ciccarelli, T.F. Quatieri, and S.S. Ghosh, "Neurophysiological Vocal Source Modeling for Biomarkers of Disease," paper in *Proceedings of INTERSPEECH 2016: Understanding Speech Processing in Humans and Machines*, 2016, pp. 1200–1204.

50. F.H. Guenther, S.S. Ghosh, and J.A. Tourville, "Neural Modeling and Imaging of the Cortical Interactions Underlying Syllable Production," *Brain and Language*, vol. 96, no. 3, 2006, pp. 280–301.

51. M. Valstar, B. Schuller, K. Smith, F. Eyben, B. Jiang, S. Bilakhia, et al., "AVEC 2013: The Continuous Audio/Visual Emotion and Depression Recognition Challenge," paper in *Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge*, 2013, pp. 3–10.

52. B.M. Bot, C. Suver, E.C. Neto, M. Kellen, A. Klein, C. Bare, et al., "The mPower Study, Parkinson Disease Mobile Data Collected Using ResearchKit," *Scientific Data*, vol. 3, 2016, article no. 160011.

53. J.R. Williamson, D.W. Bliss, D.W. Browne, and J.T. Narayanan, "Seizure Prediction Using EEG Spatiotemporal Correlation Structure," *Epilepsy & Behavior*, vol. 25, no. 2, 2012, pp. 230–238.

About the Authors

Thomas F. Quatieri, a senior member of the technical staff in the Human Health and Performance Systems Group, is involved in bridging human language and bioengineering research and technologies. Within this group, he has initiated and developed major R&D and technology transition programs in speech and auditory signal processing and neuro-biophysical modeling with application to detection and monitoring of neurological, neurotraumatic, and stress conditions. He has been an author on more than 200 publications, holds 11 patents, and authored the textbook *Discrete-Time Speech Signal Processing: Principles and Practice*. He also holds a faculty appointment in the Harvard-MIT Speech and Hearing Bioscience and Technology Program. He developed the MIT graduate course Digital Speech Processing and is active in advising graduate students on the MIT and Harvard campuses. He is a recipient of four IEEE Transactions Best Paper Awards and the 2010 MIT Lincoln Laboratory Best Paper Award. He led the Lincoln Laboratory team that won the 2013 and 2014 AVEC Depression Challenges and the 2015 MIT Lincoln Laboratory Team Award for their work on vocal and facial biomarkers. He has served on the IEEE Digital Signal Processing Technical Committee, the IEEE Speech and Language Technical Committee, and the IEEE James L. Flanagan Speech and Audio Awards Committee. He has served on National Institutes of Health and National Science Foundation

panels, been an associate editor for the *IEEE Transactions on Signal Processing*, and is an associate editor of *Computer, Speech, and Language*. He has been an invited speaker at conferences, workshops, and meetings, most recently as a keynote speaker at the 2019 Speech Science and Technology Conference. He is a Fellow of the IEEE and a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, International Speech and Communication Association, Society for Neuroscience, Association for Research in Otolaryngology, and the Acoustical Society of America. He holds a bachelor's degree (summa cum laude) from Tufts University, and master's, electrical engineering, and doctoral degrees from MIT.

Laura J. Brattain works at the intersection of biomedical image processing, artificial intelligence, high-performance computing, and biorobotics. She is a technical staff member in the Human Health and Performance Systems Group and the principal investigator of a number of biomedical image processing programs. She collaborates closely with MIT campus and Boston-area hospitals, including Massachusetts General Hospital. She holds a doctoral degree in bioengineering from Harvard University.

Todd A. Thorsen is a technical staff member in the Biological and Chemical Technologies Group, where he pursues research in microfluidics-based engineering, including the design and fabrication of state-of-the-art programmable microfluidic devices, exploitation of the microscale physical properties of fluids (thermal, chemical, optical, and electrical), and the development of microfluidic platforms for medicine. Prior to joining Lincoln Laboratory, he was a faculty member in the MIT Department of Mechanical Engineering. He has authored or coauthored 34 papers and two book chapters, has 40 issued and pending patents in the field of microfluidics, and has more than 8,800 citations of his published works. He holds a bachelor's degree in biology from the University of California, San Diego; a master's degree in infectious disease from the University of California, Berkeley; and a doctoral degree in biochemistry and molecular biophysics from the California Institute of Technology.

Gregory A. Ciccarelli is a technical staff member in the Human Health and Performance Systems Group. He works on neurocomputational modeling, using model-based approaches to study depression, Parkinson's disease, traumatic brain injury, and auditory health. He holds a doctoral degree from MIT; his dissertation characterized why speech is a good biomarker of neuropsychological disorders.

Shaun R. Berry is a technical staff member in the Advanced Materials and Microsystems Group. He has worked on a variety of microelectromechanical systems and microfluidic projects, which have included developing microelectromechanical microswitches, variable-focus liquid lenses, micropumps, and non-mechanical beam-steering devices. He is currently interested in developing microsystems for active optical devices, communications, and microhydraulics. He holds a bachelor's degree in mechanical engineering from Northeastern University and master's and doctoral degrees in mechanical engineering from Tufts University. His doctoral research was focused on microfluidics and the electrowetting phenomenon.

Jeffrey S. Palmer is the leader of the Human Health and Performance Systems Group at Lincoln Laboratory. He has oversight of multiple research programs that focus on health, human performance, objective neurocognitive analytics, and biosensing via wearable, ingestible, and implantable devices. In 2010, he helped to create the first Army War College Fellowship at Lincoln Laboratory and the MIT Security Studies Program. He has given presentations at international conferences and authored book chapters and technical articles on DNA biometrics and forensics, biomechanics, cell biology, materials science, soldier nanotechnology, biological-chemical defense, polymer science, high-energy lasers, microelectronics packaging, wearable biomedical sensing in extreme environments, and neurocognitive technologies. He has served on editorial boards for journals in biomechanics, molecular science, biomedical informatics, and biosensors. He has chaired technical conferences for the National Science Foundation, Department of Homeland Security, and the IEEE. Currently, he is the vice chair (and chair-elect) of the IEEE Engineering in Medicine and Biology Society's (EMBS) Technical Committee on Wearable Biomedical Sensors and Systems and the EMBS conference editorial board for tissue engineering and biomaterials. In addition, he has served as an advisor on senior military studies of enhancing health and performance, and led a multi-agency U.S. government effort to develop automated rapid

THOMAS F. QUATIERI, LAURA J. BRATTAIN, TODD A. THORSEN, GREGORY A. CICCARELLI, SHAUN R. BERRY, JEFFREY S. PALMER,

SATRAJIT S. GHOSH, AND KWANGHUN CHUNG

human DNA analysis capabilities for field biometrics and forensics applications. Prior to working at Lincoln Laboratory, he worked at research laboratories at IBM and GE, and at the Physical Sciences Laboratory at New Mexico State University. He holds a bachelor's degree with a minor in mathematics from New Mexico State University, a master's degree from Rensselaer Polytechnic Institute, and a doctorate with a minor in bioengineering from MIT, all with majors in mechanical engineering.

Satrajit Ghosh is a principal research scientist at the McGovern Institute for Brain Research at MIT and an assistant professor in the Department of Otolaryngology at Harvard Medical School. He is also the director of the Data Models and Integration project of ReproNim, a National Institutes of Health P41 Center for Reproducible Neuroimaging Computation. His research interests span computer science and neuroimaging, specifically in the areas of applied machine learning, software engineering, and applications of neuroimaging. The primary focus of his research group is to develop platforms for knowledge discovery by integrating a set of multidisciplinary projects that span precision medicine in mental health, imaging genetics, machine learning, and dataflow systems for reproducible research. He is a lead architect of the Nipype dataflow platform; an ardent proponent of decentralized and distributed web solutions for data sharing, querying, and computing; and a strong believer in solving problems through collaboration and crowdsourcing. To help accelerate building an open Commons for neuroscience, he is currently serving as a co-editor-in-chief for a new journal, *BMC NeuroCommons*.

Kwanghun Chung is an assistant professor of chemical engineering at the MIT Institute for Medical Engineering and Science and a principal investigator at the Picower Institute for Learning and Memory. He is a co-inventor of a novel technology termed CLARITY, which enables system-wide structural and molecular analysis of large-scale intact biological samples, including rodent brains and human clinical samples. His research laboratory at MIT is an interdisciplinary research team devoted to developing and applying novel technologies for integrative and comprehensive understanding of large-scale, complex biological systems. He holds a doctoral degree from the Georgia Institute of Technology and completed his postdoctoral training at the Karl Deisseroth Lab at Stanford University.

Open Body Area Network Physiological Status Monitor

Brian A. Telfer, Kate Byrd, and Paula P. Collins

Lincoln Laboratory has developed and successfully transitioned to industry a prototype wearable sensor that monitors the thermal-work strain of an individual. Heart-rate data are processed on the device to estimate core body temperature; skin temperature and motion data are also collected. This information is used to produce a simple real-time actionable alert on the individual's physiological status that is then sent to an Android smartphone, allowing one person to assess the physical well-being of all members in a group. Designed for use by small military units during operational or training exercises, the networked monitors allow unit leaders and medics to make informed data-driven decisions that assure service members' health and performance are not compromised.

>> **Dismounted warfighters in austere** environments face numerous physical, cognitive, and psychological stressors that reduce performance and lead to injury. Real-time physiological status monitors (RT-PSMs) can provide actionable information to small-unit leaders and medics to enable data-driven decisions that help them to maintain and improve warfighter performance and safety. The U.S. Army Medical Research and Materiel Command and its subordinate laboratory, the U.S. Army Research Institute of Environmental Medicine (USARIEM), have researched RT-PSMs through numerous technology iterations for the past two decades [1]. MIT Lincoln Laboratory first supported this effort in the early 2000s by conducting a study for USARIEM to develop an RT-PSM concept based on technology at that time [2].

Since about 2009, the year of the first Fitbit, decreases in the size, weight, and power (SWaP) of commercial electronics have led to a wave of consumer wearable sensors sold by hundreds of companies with billions of dollars in sales. It would be desirable if the military could simply purchase these low-cost commercial devices. However, with few exceptions, these commercial devices do not provide the information that small-unit leaders can act on. For example, heart rate, which is of high interest for consumers, does not in itself provide actionable information to a leader and is more likely to result in distraction. Integrating heart rate and other sensor measurements, such as skin temperature, into an individualized physiological model can provide clear actionable information, such as an alert that specific team members need to cool off. However, consumer-grade
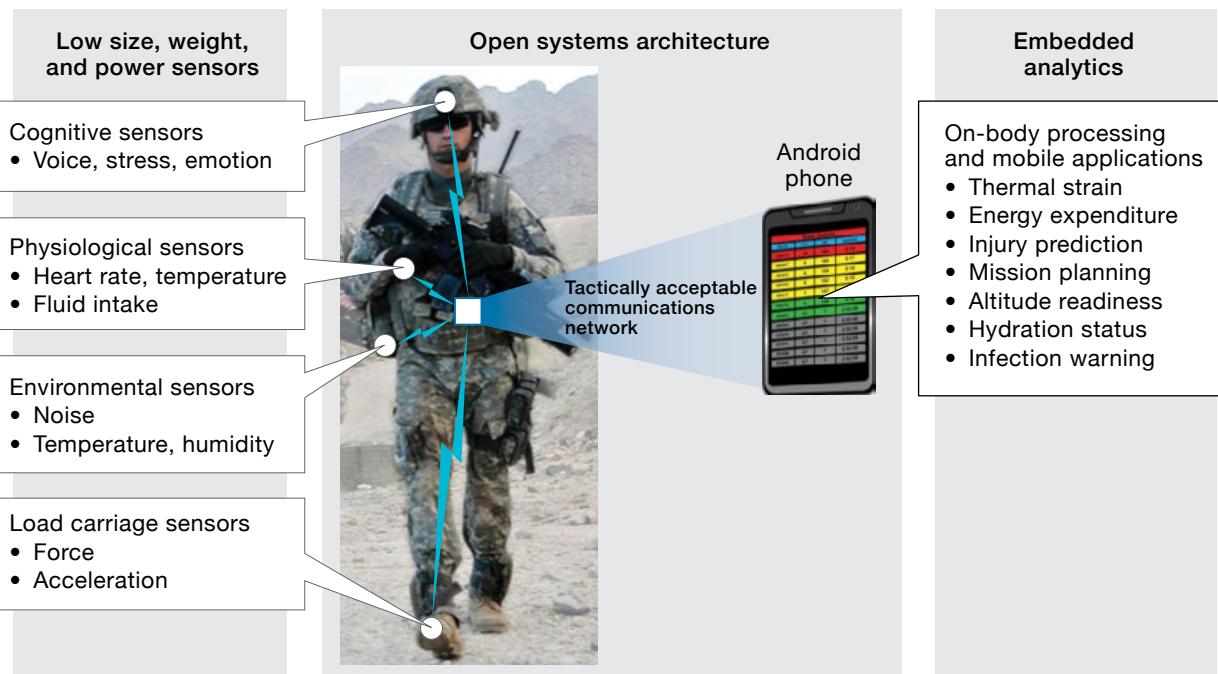
FIGURE 1. The main components of a real-time physiological status monitor (RT-PSM) are biomedical sensors, an architecture that can accommodate new technologies, and predictive analytics for providing actionable information to users. Lincoln Laboratory's experience with low-size, -weight, and -power (SWaP) sensors, open systems architectures, and embedded computing is contributing to the military's development of RT-PSMs that provide relevant evaluations of a warfighter's health and performance status while being acceptable for use in tactical and training environments.

devices are often proprietary closed systems, meaning that raw sensor data cannot be accessed for use in military-relevant algorithms, and predictive models specific to military needs cannot be integrated into their applications (apps). Other features of consumer devices—for example, relatively short battery life and reliance on wireless communications—make them insufficient or unacceptable for military use in tactical environments.

Since 2012, Lincoln Laboratory has worked with USARIEM, the Marine Expeditionary Rifle Squad (MERS), and several other government organizations to develop a series of RT-PSM prototypes to meet key requirements. As highlighted in Figure 1, these requirements include low-SWaP sensors, an open systems architecture that allows the integration of new sensors, analytics to generate real-time actionable information from sensor measurements, and tactically acceptable communications. RT-PSM prototypes have included the Open Body Area Network Physiological Status Monitor (OBAN PSM) to address thermal work strain, the Mobility and Biomechanics Insert for Load Evaluation (MoBILE)

to address musculoskeletal overuse injury, and a tactical noise dosimeter to address hearing damage from operational exposures to extreme noise levels.

The OBAN PSM is the most mature of the RT-PSM efforts. The Lincoln Laboratory prototype technology was transitioned via a competitive bid process to industry to mature into a product. The winning contractor, Odic Inc., successfully completed the delivery of 300 units for use by the Army and Marines in 2018 field tests. Additional improvements are underway to support system employment at scale in military training environments, with the goal of establishing a leave-behind safety monitoring capability at Fort Benning, Georgia. The timeline for the development, testing, and transition of the OBAN PSM prototype is provided in Figure 2.

Physiological Model

The Physiological Strain Index (PSI), also referred to as Heat Strain Index (HSI) [3, 4], is a tool to identify overheating in individuals. Overheating occurs when a body cannot maintain core temperature (TC) at a safe

level. Hot and humid conditions, especially when coupled with personal protective equipment, make it hard for the body to cool itself efficiently. Warfighters, first responders, and chemical, biological, radiological, and nuclear (CBRN) defense personnel are particularly prone to heat injury [3, 5–8].

Thermal-work strain is a measure of overheating (thermal) during periods of activity (work). High levels of thermal-work strain can cause heat injury at two levels. The first level is heat exhaustion, whose symptoms may include dizziness, thirst, weakness, and headache [9, 10]. If high levels of thermal-work strain persist, they may lead to heat stroke, a life-threatening condition that can cause delirium, convulsions, or coma [9, 10]. Even short of reaching heat exhaustion, thermal-work strain (i.e., heat strain) will degrade mental and physical performance capabilities [11], reducing a warfighter's combat readiness and increasing the likelihood of other injuries.

The PSI was developed by the USARIEM in response to the calculation complexity and inflexibility of the models and indices that preceded it [12]. As stated by Moran, the PSI "should be capable of indicating heat strain…and is expected to be sensitive enough to differentiate between similar exposures that differ in one variable," such as exposures which vary only in metabolic rate, level of protective clothing, or climate [12]. Since the PSI's development, Moran and others have validated it across hydration levels [13], genders [14], clothing levels [15], environmental conditions [14, 15], metabolic workloads [14–17], aerobic capacities [18], and ages [18]. Of the 14 thermal strain indices identified by de Freitas and Grigorieva [19], PSI remains the only one that operates across the desired range of conditions whose inputs are measurable from the body in real time.

The PSI combines information on thermal burden and work strain to provide a single, simple 0–10+ scale to assess physiological status [20]. The thermal and workload portions of the PSI are computed from core body temperature and heart rate, respectively. When core temperature rises from the normal 37 degrees Celsius up
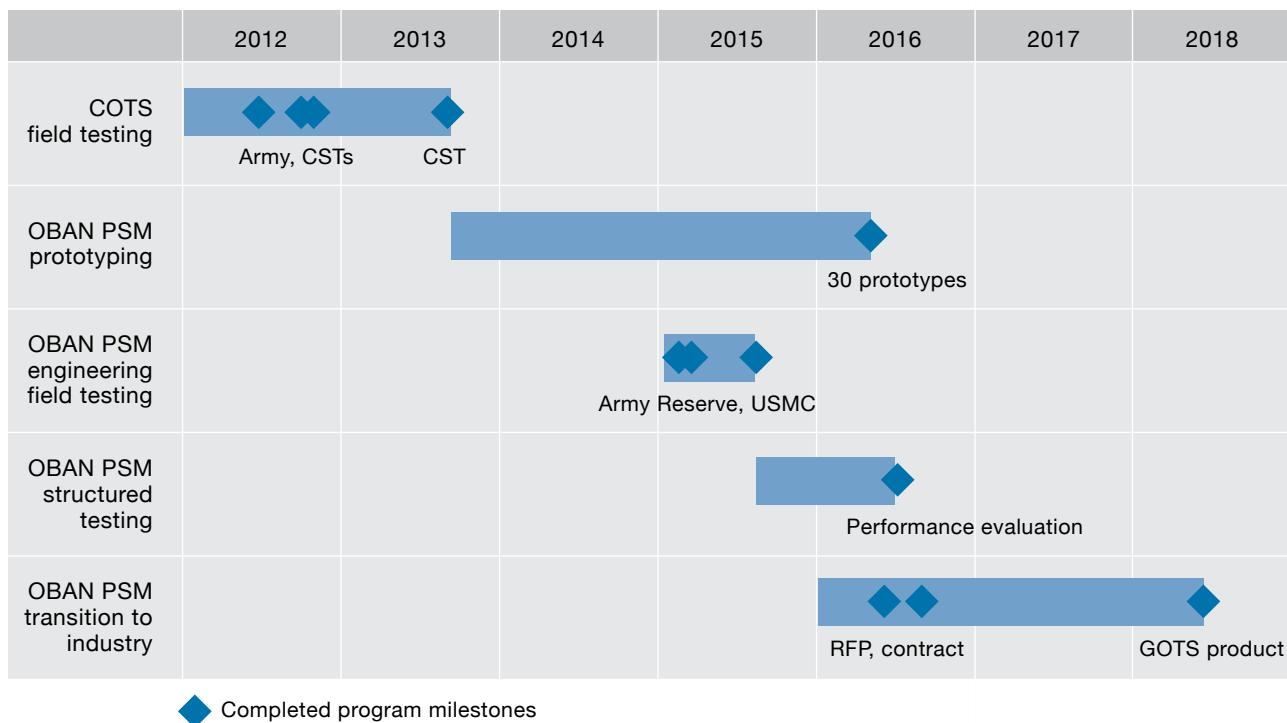
FIGURE 2. The development, testing, and transition timeline for the Open Body Area Network Physiological Status Monitor (OBAN PSM) illustrates the stages of the project spanning more than six years. Prior to the start of Lincoln Laboratory development, commercial devices were evaluated with military end users for suitability. Their identified shortcomings necessitated the government investment into the OBAN PSM development effort. COTS: commercial off-the-shelf. CST: National Guard Civil Support Team. RFP: request for proposals. GOTS: government off-the-shelf.

to 40 degrees Celsius or higher during prolonged exertion, heat stroke can occur. Heat exhaustion and heat stroke are known to occur in both training and combat, but are commonly underreported. One well-publicized example occurred at Fort Bragg, North Carolina, in September 2011 when 43 out of 60 soldiers experienced heat injuries during a 12-mile road march. Eighteen of the soldiers were taken to the hospital, and one was admitted to the intensive care unit [21].

The gold standard for measuring core temperature in the field is an ingestible temperature capsule, which is effective as a research tool but is deemed unacceptable for training and operations. To address this deficiency, USARIEM developed and validated an algorithm for estimating core temperature. The initial model is based on smoothed heart-rate estimates. The physiological basis for this algorithm is twofold. First, heart rate increases with work, which heats the body core. Muscles are only about 20 percent efficient, with 80 percent of the energy generated during work going to heat production [22]. Second, heart rate increases to support the body's heat dissipation. To dissipate heat, blood vessels near the skin vasodilate to increase blood perfusion. Thus, heart rate increases both to support the cardiac output needed to perform work and to increase skin blood flow to allow dissipation of the resulting heat.

Estimated core temperature can thus serve as an input into the PSI model, allowing thermal-work strain to be predicted from time-series heart-rate data alone. However, anticipated model improvements incorporate skin temperature measurements to help determine the level of heat strain compensation provided by evaporative cooling [23]. Activity level and neuromotor coordination measured from accelerometry may also be important inputs toward creating a more actionable alerting capability. A key goal of the OBAN PSM is to bring together these three real-time measurements—heart rate, skin temperature, and accelerometry—into an integrated platform that is acceptable for military use.

System Overview

The OBAN PSM was initially designed to monitor a team of up to 13 members. The number of team members was chosen to accommodate both Army and Marine squads, currently sized at 9 and 13, respectively. The team leader or medic monitors physiological status on an Android
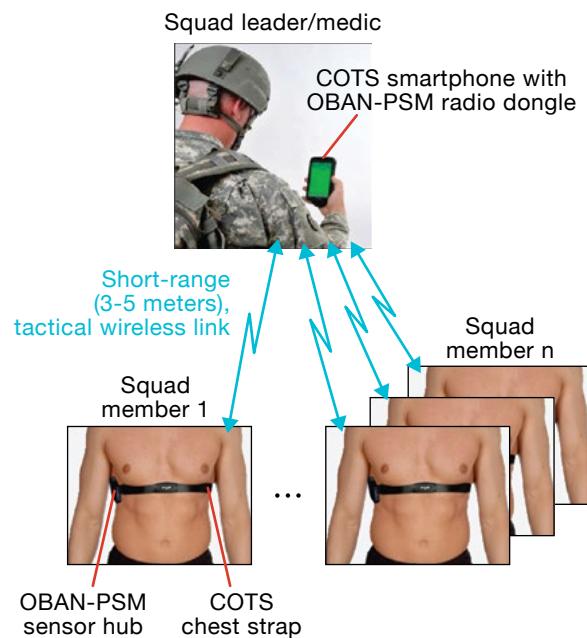
FIGURE 3. The OBAN PSM system allows team leaders or medics to monitor the current physiological status of their squads via a short-range tactical wireless link.

smartphone, referred to in the Army as an end-user device (EUD), via a short-range (3–5 meters) tactically acceptable OBAN radio. This short-range body area network allows for updates during the regular, up-close status checks that a leader normally conducts with his or her team members. Alternatively, the OBAN can provide a wireless, on-body link to a longer-range radio worn elsewhere on body. The operating concept is depicted in Figure 3.

Leveraging commercial technology, Lincoln Laboratory researchers developed a prototype on-body device. Starting with a Polar chest strap and sensor (Polar Electro, Kempele, Finland), the researchers designed a "hub" to contain the required additional sensing, computing, and wireless communications electronics. Figure 4 depicts the components of the prototype OBAN PSM. The Polar sensor detects heartbeats from cardiac electrical activity and communicates beat times over a short-range (1 meter) inductive wireless link to a Polar receiver in the hub. The hub then computes heart rate and PSI. It also senses skin temperature and accelerometry, logs data, and wirelessly communicates heart rate and PSI off-body. The Polar sensor is attached to the chest strap that is worn on the front of the body centered over the

heart, while the hub is worn on the side of the chest to be compatible with body armor.

The EUD also heavily leveraged commercial technology in the form of a Samsung Note 3 Android phone. The EUD includes an OBAN dongle and whip antenna, an extended-life battery, and a case. Cellular communications are disabled. The OBAN PSM EUD is a surrogate for the Army's Nett Warrior EUD, which also leverages commercial Android cell phone technology. Figure 5 depicts the electronics, packaging, and display software for the EUD.

Key needs for the wireless body area network technology are (1) minimum power to prolong battery life while transmitting infrequent, small-packet-size status updates and (2) compatibility with military wireless communications. Since Bluetooth and other
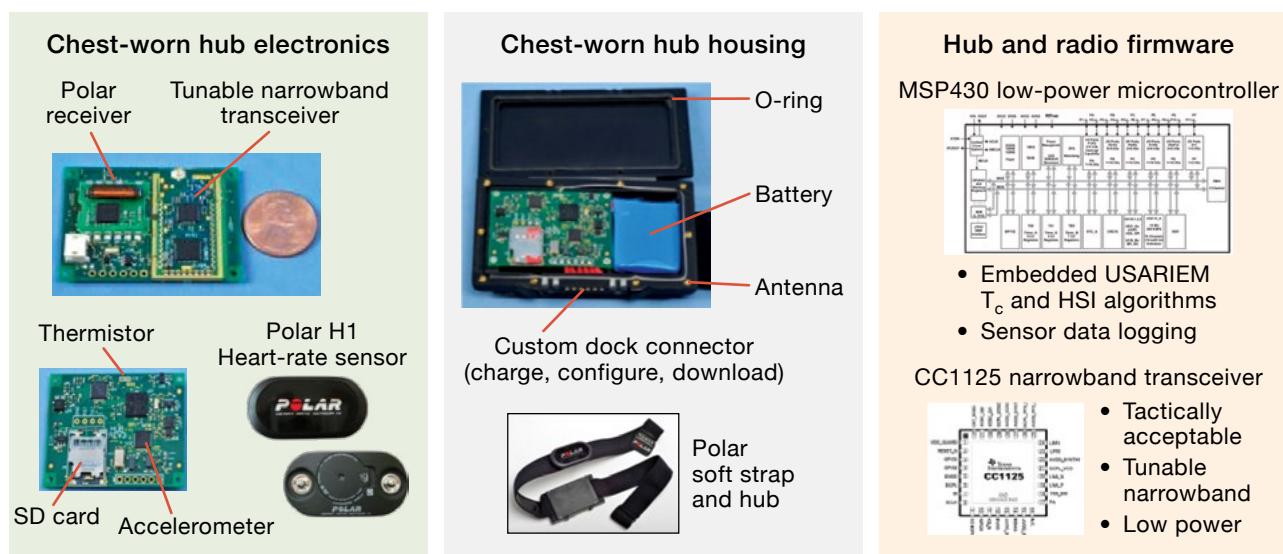
FIGURE 4. The chest-worn components of the prototype Open Body Area Network Physiological Status Monitor include the custom-built hub electronics and housing, and the embedded software that enables data analysis and communication.

FIGURE 5. The prototype end-user device for the Open Body Area Network Physiological Status Monitor uses commercial off-the-shelf (COTS) components that were adapted for use with custom software and radio hardware.

common-place commercial solutions would not be acceptable, other solutions were surveyed. The best identified option at the time (2013) was a tunable narrowband (TNB) low-power commercial transceiver operating with a messaging protocol designed by Lincoln Laboratory staff. The TNB radio is implemented on a daughterboard that is mounted on the hub and EUD dongle circuit boards. The EUD transmits a query for updates, triggering hubs within range of the EUD to transmit heart rate and PSI values. The radio communicates over a standard narrowband military frequency channel, which is assigned during the configuration process. The TNB radio has been approved by the U.S. Army through the DD1494 process for frequency allocation at military installations and operating environments.

To complete the system, a configuration laptop and docking station are required for charging and downloading data.

Engineering Field Tests

After an initial build of nine prototype devices, three engineering field tests were conducted by Lincoln Laboratory and USARIEM with the goals of identifying any technical performance issues in relevant environments and obtaining warfighter feedback on the concept of employment and human factors. The testing team was successful in obtaining feedback, in marking general successes and failures, and in identifying several system issues. A system failure analysis was performed to determine hardware and software weaknesses and to establish corrective actions that could be applied to a 30-hub build-up. This failure analysis provided the motivation for additional rigorous structured testing.

Two engineering field tests were conducted with the Army Reserve 743rd Transportation Company at Hanscom Air Force Base (HAFB), Massachusetts, and at the Camp Ethan Allen Training Site (CEATS), Vermont; one was conducted with the U.S. Marine Corps School of Infantry–East (SOI-E), Camp Geiger, North Carolina. The volunteers, 15 men and 6 women, enrolled and participated in the tests after reading and signing an informed consent form. The tests were conducted according to procedures approved by the USARIEM Human Use Review Committee, the MIT Committee on the Use of Humans as Experimental Subjects, and the U.S. Army Medical Research and Materiel Command Human Research Protection Office. The test durations progressively increased: 3 hours for HAFB, 24 hours for CEATS, and 57 hours at SOI-E. Note that participants slept while wearing the devices during the latter two tests. Photographs from the three tests are shown in Figure 6. Three aspects of these tests are reported herein: communications performance, sensor measurement quality, and reliability.

FIGURE 6. Photographs show scenes from the engineering field tests. Clockwise from upper left are scenes from Hanscom Air Force Base, the Camp Ethan Allen Training Site, and the U.S. Marine Corps School of Infantry–East.

Communication Performance

During the engineering tests, the EUDs were set to operate in manual mode, i.e., the EUD operator pushes a button on the display to send a query. Because the tests were integrated into routine warfighter training activities, uncontrolled separations between the EUD and hub were experienced. Thus, statistics on queries and replies gave a qualitative indication of communication performance.

Quantitative range testing conducted separately under controlled conditions measured 95 percent message transmission at 2.7-meter range, with and without body armor, and independent of the subject's orientation to the EUD. Although this transmission percentage would be unacceptably low for a general-purpose radio, it is judged to be acceptable for a low-power body area network that transmits status information. If desired, the percentage could be increased by simply repeating a query when a reply is not received.

## SOI-E Sensor Measurement Examples

An important activity for evaluating the OBAN PSM as a heat strain monitor was the SOI-E 15-kilometer ruck march, which served as a final pass/fail event for training completion. As shown in Figure 7a, PSI values ranging between 0 and 10 were observed, increasing as expected over the duration of the march. The wide range of values across Marines at any particular point in time highlights the value of individualized, real-time monitoring. When the EUD display alerted support personnel that one Marine was approaching and then reached a PSI of 10 (shown in Figure 7b), extra attention was paid to ensure the safety of that Marine until the march was completed.

High-quality data were logged for heart rate, accelerometry, and skin temperature for a total of 57 hours for seven hubs. Figure 8 shows example sensor data from one volunteer collected when the volunteer was asleep. Figure 9 shows data collected when the volunteer was on the ruck march. Two issues were identified during the data collection. First, in approximately 0.3 percent of the data collected, the heart rate was measured as half of the true heart rate. An algorithmic correction for this was implemented during follow-on development. Second, 10 hours of accelerometry data were not logged for one hub (2.5 percent of the total). This was due to a software bug that was also corrected during follow-on development.

Structured Testing

After iterative testing with users was completed and improvements were implemented on the basis of lessons learned and component failure analysis, a set of 30 hubs was built. These devices were subjected to a rigorous laboratory and on-body testing protocol, referred to as structured testing. Tests evaluated sensor accuracy, battery life, radio functionality, and ability to operate in harsh environments. In addition, the 30 hubs were worn by volunteers in a variety of free-living activities. The results of this testing are summarized here:

- Heart-rate accuracy. The 95 percent confidence interval of errors, based on human subject data, was measured as [−3.2, 1.3] beats per minute over a range of activities. This accuracy is consistent with that of the Polar heart-rate sensor that is used by the OBAN PSM prototype.

- Accelerometer accuracy. Accelerations for the three axes were found to fall within 10 percent of those from a truth sensor, measured on a vibration table with peak acceleration of 6g. Accuracy was higher for lower levels of acceleration. Noise levels were higher than those from a truth sensor. This high noise level was determined to be caused by insufficient voltage regulation and was addressed in the technology transition.

- Thermistor accuracy. After applying a sensor-specific calibration offset, all 30 hubs were measured to be accurate to within 0.2 degrees Celsius over the range of 30 to 47 degrees Celsius. The thermistors' dynamic response was measured and was determined to be sufficiently fast to track skin temperature.

- Core temperature estimation and PSI algorithm implementation. These implementations were anchored against reference implementations and found to perform identically.

- Wireless performance. Wireless communication was tested in the lab (off-body) at 1-meter range with more than 100,000 EUD requests and more than 1,000,000 hub replies during upwards of 1,100 hours. Median message loss was 0.28 percent, and 78 percent of the tests yielded loss rates of less than 0.5 percent. Message loss measured on-body at three yards (2.7 meters) was 5 percent, with or without body armor, and with no dependence seen on the side of the body facing the EUD. An interference test was conducted with three teams, each wearing 10 hubs, with each team assigned

(a)

(b)
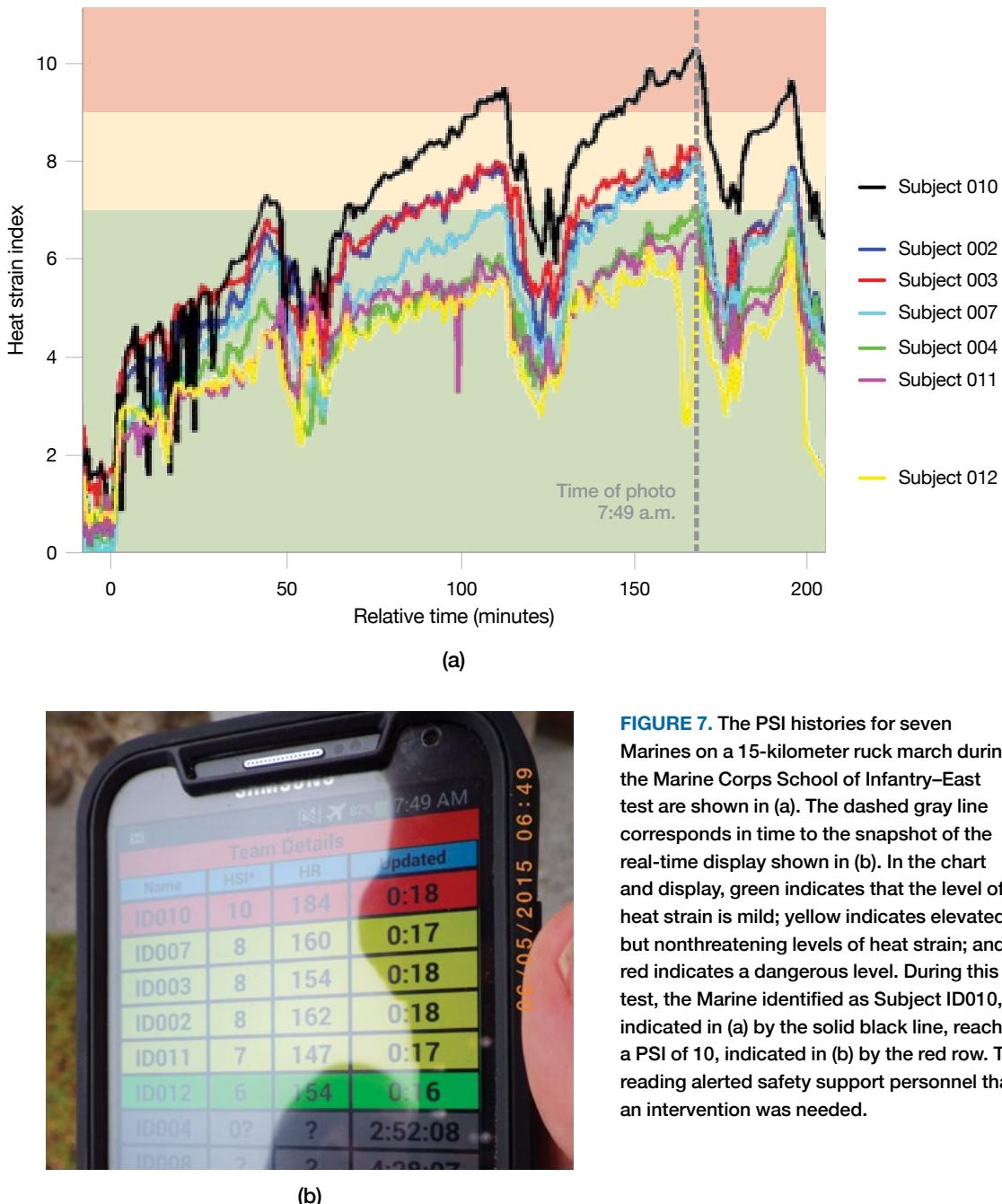
FIGURE 7. The PSI histories for seven Marines on a 15-kilometer ruck march during the Marine Corps School of Infantry–East test are shown in (a). The dashed gray line corresponds in time to the snapshot of the real-time display shown in (b). In the chart and display, green indicates that the level of heat strain is mild; yellow indicates elevated but nonthreatening levels of heat strain; and red indicates a dangerous level. During this test, the Marine identified as Subject ID010, indicated in (a) by the solid black line, reached a PSI of 10, indicated in (b) by the red row. This reading alerted safety support personnel that an intervention was needed.

a different frequency channel. There was no degradation of communications for the three teams when they were standing in close proximity in a room.

- Battery life. Six hubs were tested in the lab, resulting in measured battery lives of 4.8 to 5.5 days.
- Robustness to heat, vibration, and sweat. Temperature was varied from 13 to 55 degrees Celsius in a chamber. None of the 30 devices failed, although the message loss exceeded 1 percent for four hubs at the higher

temeratures. This temperature dependency is understood and was addressed in the technology transition. For vibration, four hubs were tested on a vibration table that ran a standard profile simulating a Humvee driving over rough terrain. No failures or anomalies were encountered. To simulate a sweaty environment, three hubs were sprayed with a saline solution and then wrapped in a saline-soaked paper towel for one hour. No failures or anomalies were encountered.
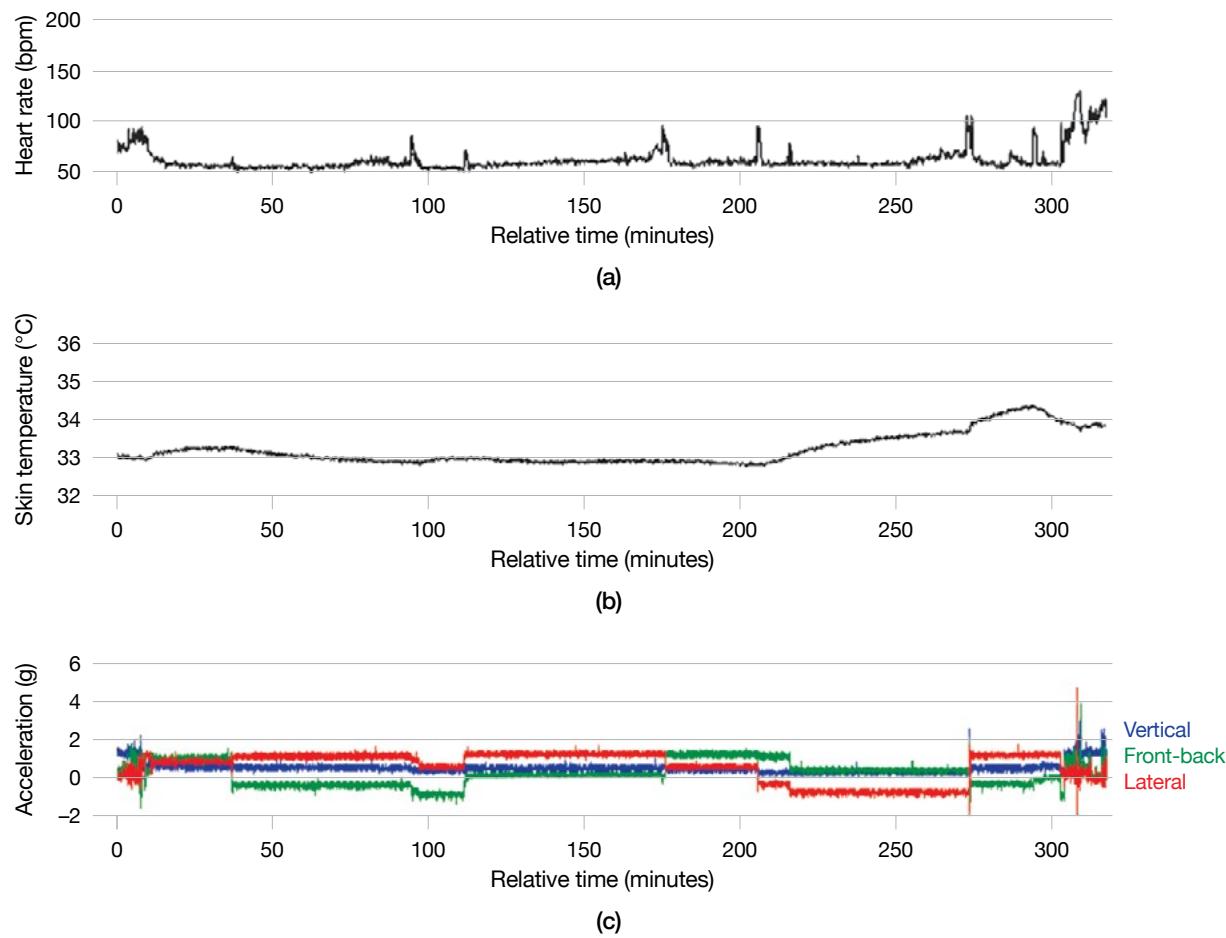
FIGURE 8. Sensor data collected during sleep, as shown here for one SOI-E Marine, can provide rich insight into sleep quality and other health factors that can impact performance during subsequent activities. Note that the accelerometry data show changes in resting position that correlate with spikes in heart-rate data.

- On-body reliability. The 30 hubs were worn repeatedly by Lincoln Laboratory staff during varied physical activities for a total of more than 180 hours. No failures were encountered.

Structured testing made use of a variety of the Laboratory's test equipment and facilities, including the RF Compact Range for radio testing, environmental chambers, and mechanical test equipment. As an example, vibration exposure was performed to evaluate whether rigorous travel conditions and activity could be expected to cause device failure in the OBAN hubs. For this test, four hubs were mounted to a single-axis T2000 vibration table (Unholtz-Dickie, Wallingford, Connecticut) with fixtures designed in-house for this purpose; only four hubs were used because of space limitations on the vibration table. A Polar heart-rate simulator was positioned alongside the vibration plate and a Polar watch (FT1 or RS100) was used to check the heart-rate truth at the test's start and finish. The EUD was set up approximately two meters from the vibration table shown in Figure 10 and set to query for updates every 30 seconds. To capture a baseline for this new testing environment, the EUD started to query for updates from the hubs 1.5 hours before the hubs were exposed to vibrations. An operationally representative Humvee profile was selected for this test. It ranges in frequency from 5 to 500 Hz and in magnitude from −8g to +8g. The profile was run for 30 minutes in each axis. No failures or anomalies were noted.

Technology Transition

After a competitive bidding process, Odic Inc. (Devens, Massachusetts) was selected to mature the OBAN
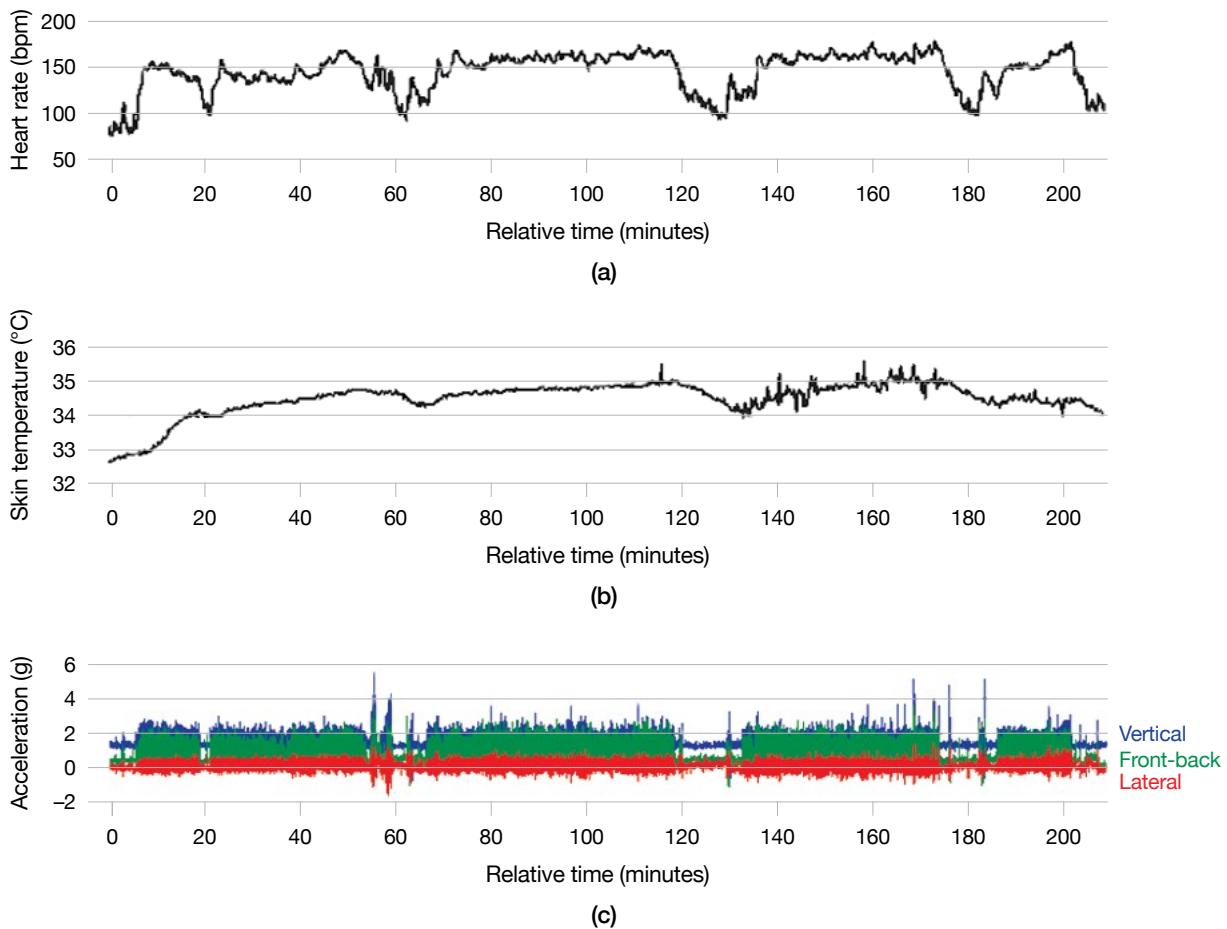
FIGURE 9. Sensor data collected from one SOI-E Marine during the 3.5-hour ruck march show the high level of sustained activity measured by the accelerometer and the resulting elevated heart rate. Recovery during the four rest breaks, noted by the periods of low acceleration measurements, can also be seen.

PSM prototype into a product. The result is shown in Figure 11. The system includes significant improvements and innovations. The single-module on-body device is significantly smaller, and it replaces the Polar sensor by snapping directly into the commercial off-the-shelf chest strap. Battery life has been extended to one week, and can be extended longer if data logging is disabled. In addition to the TNB transceiver, Bluetooth Low Energy, a variant of Bluetooth wireless technology, has been added as an option. Bluetooth Low Energy is convenient for non-tactical use, such as in domestic military training environments. The functionality of the phone dongle and the antenna have been integrated into the Moto mod add-on, which snaps onto the Moto Z Android phone, to produce a single integrated unit. The docking station is a single, compact unit that charges and

downloads data from the on-body devices. As a result of this project's success, Odic was named the U.S. Small Business Administration's Subcontractor of the Year for the New England Region. Significant portions of Lincoln Laboratory's prototype have been extended and included in the product, including much of the firmware, the phone app, and the configuration PC software.

The system, consisting of 300 on-body devices, 20 EUDs, and 20 docking stations, was delivered by Odic to the Laboratory in May 2018. The devices were then loaned to the Army and Marines for field testing. Approximately 2,000 service members wore the OBAN PSM devices in 2018 and 2019 during training programs at multiple hot locations, primarily at Fort Benning, Georgia, under a human subject research protocol conducted by USARIEM. The collected data are currently

undergoing analysis to further refine the PSI algorithm. Additional field testing was conducted by USARIEM to assess the usability of the system by military personnel encapsulated in the protective gear required for chemical and biological threat response. An additional round of system improvements is currently underway by Odic and Lincoln Laboratory, to be implemented prior to large-scale purchase by the military.

Future Directions

The OBAN PSM system is a significant advance in wearable monitoring for the military. The system is suitable for operational use because of its ability to provide actionable information via a wireless body area network in a small device with a one-week battery life. Although the cost in volume remains to be determined, the system's use of low-cost commercial components and design for manufacture are expected to result in an affordable cost. This affordability will allow the government to scale up the use of real-time thermal-work strain monitoring to first reduce the number of heat injuries in training environments. As the number of users increases, their collected data will allow the physiological models and actionable alerting to be improved, further extending the capabilities and employment of physiological monitoring. The system's open, government-owned architecture will allow additional wearable technologies to be integrated as they mature.

The use of a chest strap to measure heart rate is perhaps technically mundane but is an important consideration for comfort. Acceptability of chest straps has been found to vary from individual to individual, and acceptability for 24 hours may not extend to a week of wear. The objective solution is to integrate long-lived real-time monitoring within clothing and equipment. For example, commercial compression shirts are available to measure heart rate; however, individual variation in electrode fit has been an issue. More advanced methods for shirt integration, such as functional fibers that can themselves serve as sensors, are being investigated by the Army. Epidermal electronics or removable tattoos may prove suitable, although the Army has found in the past that adhesive patches are not acceptable to soldiers in hot, dirty, sweaty environments. Regardless of the particular type of integration with the body in the future, the government-owned open architecture body area network will evolve and continue to be relevant.
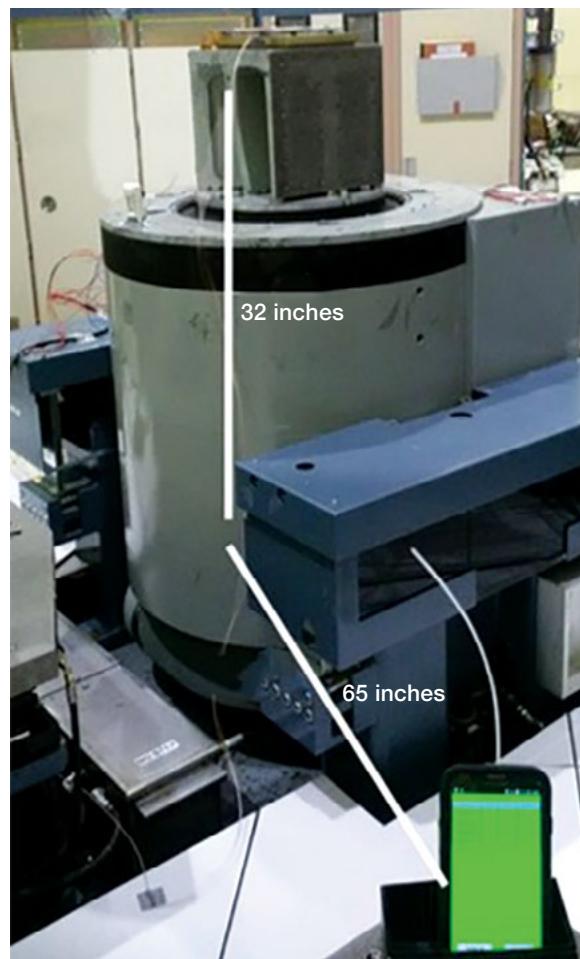
FIGURE 10. For the *z*-axis vibration durability test, the cell phone (EUD) was set up 65 inches from the vibration table on which the hubs were positioned. Additional dimensions are provided for scale. A Humvee vibration profile was selected for this test as representative of stressing field conditions.

Beyond thermal-work strain, the government has identified the need to concurrently measure multiple dimensions of human readiness, including cognitive, musculoskeletal, and immune system status, as well as environmental exposure, as was shown in Figure 1 [1]. The capability to monitor these aspects of readiness with wearable sensors is under development, but these sensors are expected to require additional sensing nodes at the head and feet. These additional sensing nodes will also need to be integrated within the open architecture.

As RT-PSM becomes integrated into equipment, decreasing SWaP will continue to be important. PsiKick (Charlottesville, Virginia) tailored its development of an ultralow-power system on chip (SoC) for the Internet
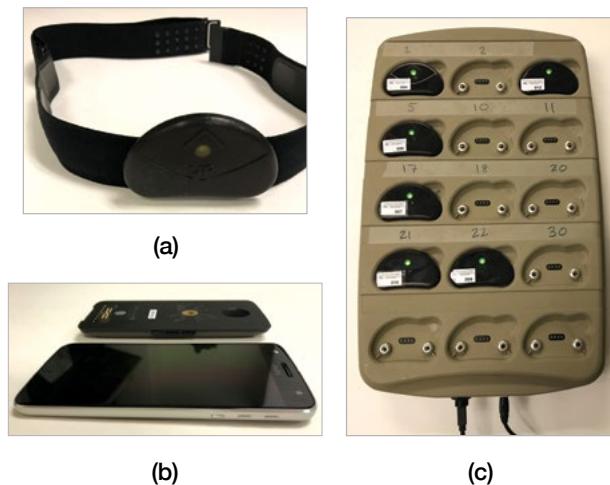
FIGURE 11. The Open Body Area Network Physiological Status Monitor product is composed of an on-body device (a), the Moto Z Android phone with Moto mod add-on (b), and a docking station (c). PHOTOGRAPHS COURTESY OF ODIC INC.

of Things to address government needs for wearable sensing. This work was funded through a Lincoln Laboratory contract and the government's Small Business Innovation Research program, with additional research funded through the National Science Foundation's ASSIST (Advanced Self-Powered Systems for Integrated Sensors and Technologies) Center. A highly innovative aspect of this work was an ultralow SWaP ultrawideband transmitter, combined with a narrowband wakeup radio. Other companies are also progressing with reduced SWaP SOCs. The SOC technology is based on the same type of subthreshold logic that Lincoln Laboratory has been pioneering for years for other government applications. The goal is an SoC that consumes less than 50 microwatts to allow the physiological status monitor to operate continuously for months, or indefinitely with energy harvesting. At that point, wearable sensing will truly become "wear and forget."

Acknowledgments

References

1. K.E. Friedl, M.J. Buller, W.J. Tharion, A.W. Potter, G.L. Manglapus, and R.W. Hoyt, "Real-time Physiological Status Monitoring (RT-PSM): Accomplishments, Requirements, and Research Roadmap," U.S. Army Research Institute of Environmental Medicine, Natick, Mass., Technical Note TN16-02, 2016, pp. 1–68.

2. G. Shaw, A.M. Siegel, G. Zogbi, and T.P. Opar, "Warfighter Physiological and Environmental Monitoring," MIT Lincoln Laboratory, Lexington, Mass., Technical Report ESC-TR-2004-077, 1 Nov. 2004.

3. W.J. Tharion, A.W. Potter, C.M. Duhamel, A.J. Karis, M.J. Buller, and R.W. Hoyt, "Real-Time Physiological Monitoring While Encapsulated in Personal Protective Equipment," *Journal of Sport and Human Performance*, vol. 1, no. 4, 2013, pp. 14–21.

4. T.B. Hughes, A.R. Hess, J.R. Simpson, W.T. Young, S. Mullen, and W. Tharion, "Assessment of Two Data Transmission Communication Systems in Real-Time Physiological Status Monitoring Applications for Weapons of Mass Destruction Civil Support Teams (WMD-CSTs)," MIT Lincoln Laboratory, Lexington, Mass., Project Report PSM-1, 23 June 2014.

5. Army Medical Surveillance Activity, "Heat Related Injuries, U.S. Army," Armed Forces Health Surveillance Branch, Washington, D.C., *Medical Surveillance Monthly Report*, vol. 12, no. 5, 2006, pp. 2–4.

6. R.F. Fahy, and P.R. Leblanc, "Firefighter Fatalities in the United States—2005," *National Fire Protection Association Journal*, vol. 100, no. 4, 2006. pp. 50-63.

7. S.J. Baker, J. Grice, L. Roby, and C. Matthews, "Cardiorespiratory and Thermoregulatory Response of

Working in Fire-Fighter Protective Clothing in a Temperate Environment," *Ergonomics*, vol. 43, no. 9, 2000, pp. 1350–1358.

8. M.J. Buller, W.J. Tharion, C.M. Duhamel, and M. Yokota, "Real-Time Core Body Temperature Estimation from Heart Rate for First Responders Wearing Different Levels of Personal Protective Equipment," *Ergonomics*, vol. 58, no. 11, 2015, pp. 1830–1841.

9. J.L. Glazer, "Management of Heatstroke and Heat Exhaustion," *American Family Physician*, vol. 71, no. 11, 2005, pp. 2133–2140.

10. A. Bouchama and J.P. Knochel, "Heat Stroke," *The New England Journal of Medicine*, vol. 346, no. 25, 2002, pp. 1978–1988.

11. M.N. Sawka, C.B. Wenger, S.J. Montain, M.A. Kolka, B. Bettencourt, S. Flinn, et al., "Heat Stress Control and Heat Casualty Management," Technical Bulletin, TB MED 507/ AFPAM 48-152, Headquarters, Department of the Army and the Air Force, Washington, D.C., 2003.

12. D.S. Moran, A. Shitzer, and K.B. Pandolf, "A Physiological Strain Index to Evaluate Heat Stress," *The American Journal of Physiology*, vol. 275, no. 1, part 2, 1998, pp. R129–R134.

13. D.S. Moran, S.J. Montain, and K.B. Pandolf, "Evaluation of Different Levels of Hydration Using a New Physiological Strain Index," *The American Journal of Physiology*, vol. 275, no. 3, part 2, 1998, pp. R854–R860.

14. D.S. Moran, Y. Shapiro, A. Laor, S. Izraeli, and K.B. Pandolf, "Can Gender Differences during Exercise-Heat Stress Be Assessed by the Physiological Strain Index?" *The American Journal of Physiology*, vol. 276, no. 6, part 2, 1999, pp. R1798–R1804.

15. M.N. Sawka, W.A. Latzka, S.J. Montain, B.S. Cadarette, M.A. Kolka, K.K. Kranking II, and R.R. Gonzalez, "Physiologic Tolerance to Uncompensable Heat: Intermittent Exercise, Field vs Laboratory," *Medicine and Science in Sports and Exercise*, vol. 33, no. 3, 2001, pp. 422–430.

16. D.S. Moran, "Stress Evaluation by the Physiological Strain Index (PSI)," *Journal of Basic and Clinical Physiology and Pharmacology*, vol. 11, no. 4, 2000, pp. 403–423.

17. R.W. Gotshall, D.J. Dahl, and N.J. Marcus, "Evaluation of a Physiological Strain Index for Use during Intermittent Exercise in the Heat," *Journal of Exercise Physiology Online*, vol. 4, no. 3, 2001, pp. 22–29.

18. D.S. Moran, W.L. Kenney, J.M. Pierzga, and K.B. Pandolf, "Aging and Assessment of Physiological Strain during Exercise-Heat Stress," *American Journal of Physiology–Regulatory, Integrative and Comparative Physiology*, vol. 282, no. 4, 2002, pp. R1063–R1069.

19. C.R. de Freitas and E.A. Grigorieva, "A Comparison and Appraisal of a Comprehensive Range of Human Thermal Climate Indices," *International Journal of Biometeorology*, vol. 61, no. 3, 2017, pp. 487–512.

20. D.S. Moran, A. Shitzer, and K.B. Pandolf, "A Physiological Strain Index to Evaluate Heat Stress," *American Journal of Physiology*, vol. 275, no. 1, part 2, 1998, pp. R129–R134.

21. C. Boutelier, L. Bougues, and J. Timbal, "Experimental Study of Convective Heat Transfer Coefficient for the Human Body in Water," *Journal of Applied Physiology–Respiratory, Environmental and Exercise Physiology*, vol. 42, no. 1, 1977, pp. 93–100.

22. I. Astrand, "Aerobic Work Capacity in Men and Women with Special Reference to Age," *Acta Physiologica Scandinavica*. Supplementum, vol. 49, no. 169, 1960, pp. 1–92.

23. M. Sawka and A. Young, "Physiological Systems and Their Responses to Conditions of Heat and Cold," in *ACMS's Advanced Exercise Physiology*, P.A. Farrell, M.J. Joyner, and V.J. Caiozzo, eds. Baltimore: Lippincott Williams & Wilkins, 2018.

Appendix

Evaluation of Heart-Rate Accuracy of COTS Monitors

Commercial heart-rate (HR) monitors on chest straps use electrical-signal sensing technology that has matured over several decades. More recently, optical technology has been developed to measure HR. Consumer optical HR monitors can be worn on the wrist, forearm, forehead, or ear. Wrist-worn monitors in particular have become commonplace and are usually considered to be more comfortable than chest straps for long-term wear. However, the accuracy of optical HR monitors for military applications has been questioned. Lincoln Laboratory was asked by the U.S. Army Medical Materiel Development Activity's Medical Support Systems Project Management Office to assess optical monitors for accuracy of both HR measurements and the biomedical information computed from these measurements.

We chose to test seven commercial heart-rate monitors representative of technology available between spring 2015 and fall 2016. The set represents a variety of wear locations: three were designed to be worn on the wrist or forearm, and two each were to be worn on the chest and forehead. These monitors fulfilled several military-related criteria:
- They must be wearable for days in dirty, sweaty conditions, thus eliminating ones secured with adhesive patches.
- They must not impede most military activities, thus ruling out sensors that clip onto a finger.

- They must support continuous monitoring during activity, not just resting heart rate.

The devices fell into two sensing modalities: electrocardiogram (ECG) or photoplethysmography (PPG). In wearable devices, ECG uses external, dry electrodes to measure electrical signals emitted by the heart muscles. In PPG, a light-emitting diode (LED) shining into the skin measures changes in reflectance caused by the variation in capillary blood flow during the heartbeat cycle. We chose NorthEast Monitoring's DR200 three-lead Holter monitor as the criterion device.

Eighteen Lincoln Laboratory employees enrolled in this study after signing informed consent forms. The testing was conducted according to procedures approved by the MIT Committee on the Use of Humans as Experimental Subjects and the U.S. Army Medical Research and Materiel Command Human Research Protection Office. The testing protocol was based on one developed by the U.S. Army Research Institute for Environmental Medicine for a previous evaluation of HR monitors. Volunteers engaged in approximately 50 minutes of low-, medium-, and high-intensity activities and unstructured rest breaks. Figure A1 shows the schedule of activities. Also shown is the heart-rate data for a single subject wearing a gold-standard device along with one wrist-worn and one headband HR monitor. The
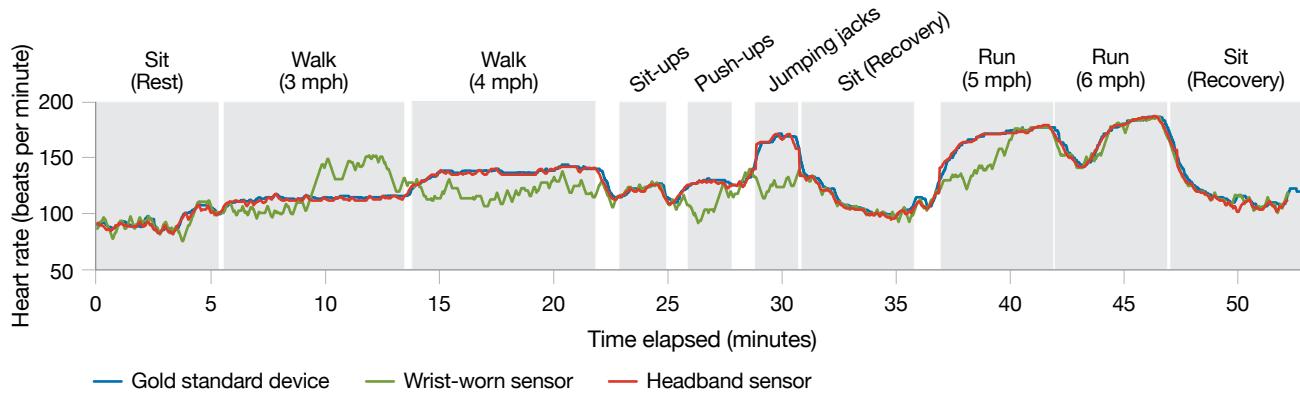
FIGURE A1. Heart-rate (HR) data are shown from a single subject concurrently wearing the criterion device and two test devices. Regions of activity are shown in gray boxes. Agreement is obtained across activity levels between the gold-standard device and the headband HR monitor, while the wrist-worn HR monitor does not track as closely.
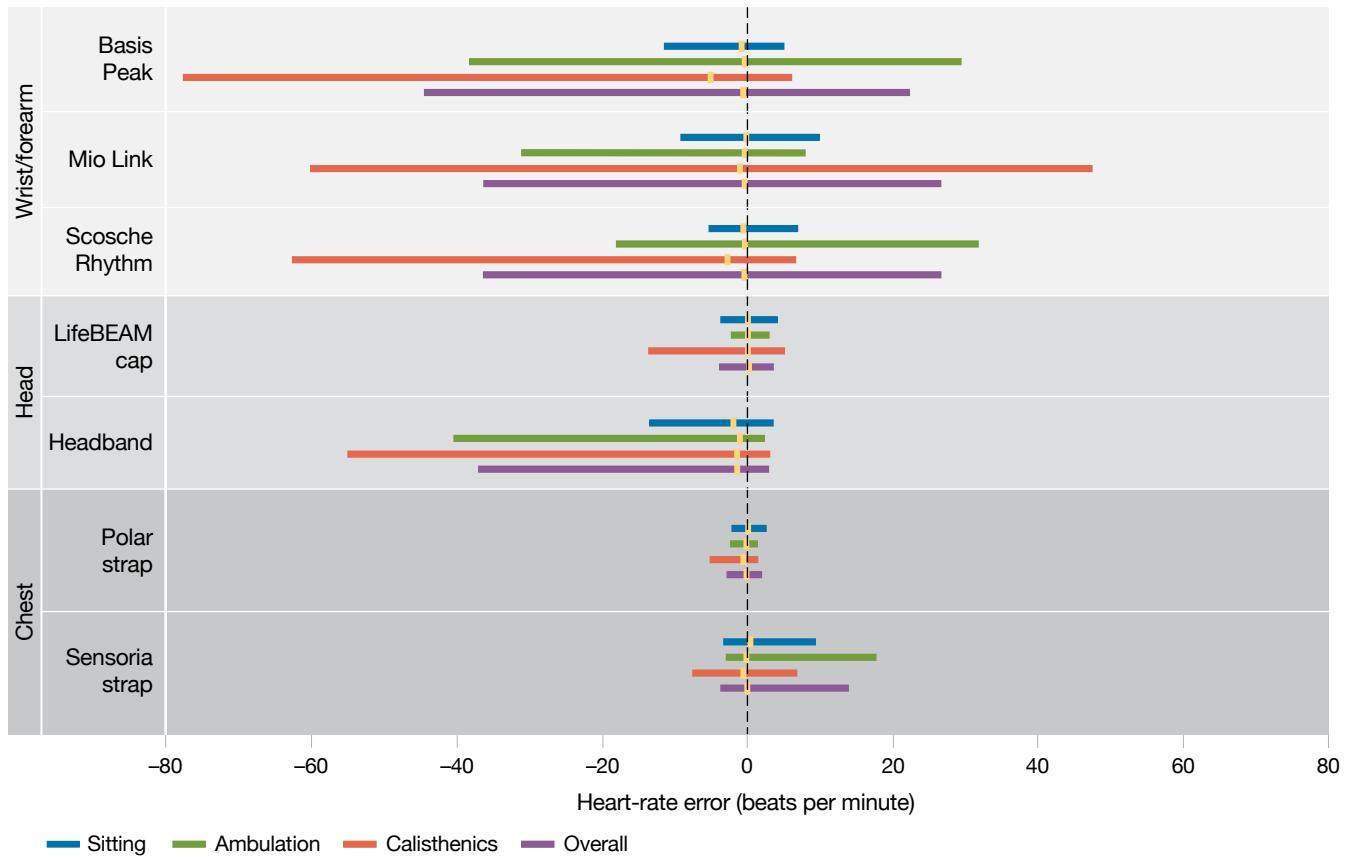
FIGURE A2. Commercial device heart-rate accuracy is shown as a function of activity and as an "Overall" metric. "Sitting" includes three dedicated periods of sitting. "Ambulation" includes two speeds of walking and two speeds of running. Sit-ups, push-ups, and jumping jacks are included in "Calisthenics." For each category, the 95 percent confidence interval is represented by the colored bar and the median by the orange mark. The Polar chest strap achieved the highest level of accuracy compared to the criterion device. Optical devices worn on the wrist and forearm were significantly less accurate, most notably during calisthenics.

figure shows the agreement across activity levels between the gold-standard device and the headband HR monitor, while the wrist-worn HR monitor does not track as closely.

Figure A2 summarizes the results of the testing by showing median HR errors and 95 percent nonparametric confidence intervals (CI). The results are broken down by type of activity, with positive differences representing an overestimate of HR by the test device, and negative differences representing an underestimate. The industry-leading Polar chest strap achieved the highest level of accuracy compared to the criterion device. The Sensoria Strap was found to have similar accuracy to the Polar Strap when it fit the subject, ensuring good electrode contact was maintained, but significantly worse accuracy when it did not fit.

The optical devices worn on the wrist and forearm were significantly less accurate than the Polar Strap.

These devices were most accurate when subjects were sitting and least accurate during calisthenics, which involved changes to wrist and hand configuration. The HR accuracy for these devices varied significantly across subjects. Finally, the LifeBEAM Smart Hat accuracy was comparable to that of the Polar Strap, but the Smart Hat would need to be integrated or made compatible with a helmet for military applications.

To determine the significance of heart-rate error we looked at its effect on scores produced by the physiological strain index (PSI), the actionable metric described in the main article. We developed an error model that assumed heart-rate errors have a Gaussian distribution and may be correlated over time. The Gaussian assumption is frequently not valid for the measured heart-rate errors, but it is an approximation that is intentionally not specific to any particular test device or activity and serves as a
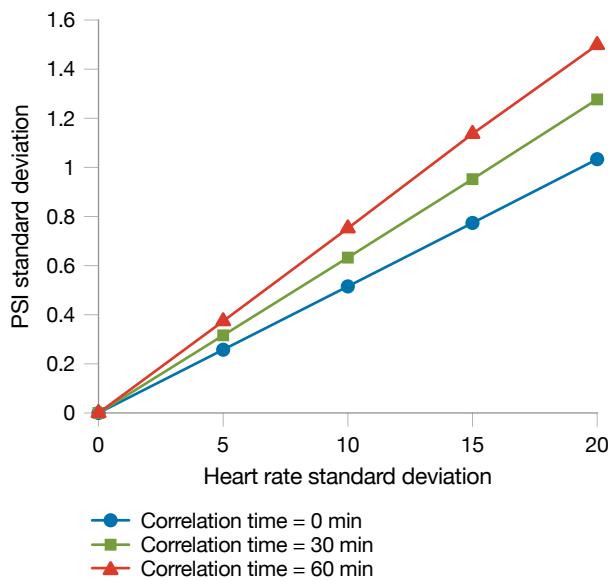
FIGURE A3. The PSI error standard deviation is plotted as a function of the heart-rate (HR) error standard deviation (beats per minute, bpm) for three HR error correlation times. It can be seen from the plot that a typical wrist-worn HR sensor's 95 percent confidence interval of greater than ±20 bpm (±2 standard deviations) results in the PSI 95 percent confidence exceeding ±1. This performance is inadequate for operational monitoring.

common point of comparison. Monte Carlo simulations were then performed with varying correlation times and HR errors, given a fixed true HR (Figure A3). If heart-rate errors are correlated over time, then a bias term magnifies the resulting error in core temperature estimates embedded in the PSI algorithm. Our error modeling led to the conclusion that the wrist- and forearm-worn device HR accuracy would be insufficient for measuring PSI to less than ±1 on a 10-point scale, and thus inadequate for operational monitoring.

Testing results from the volunteers indicated that the consumer-grade optical HR monitors were not yet sufficiently accurate for military applications. These results need to be confirmed by testing on warfighters who are conducting operational tasks. It is clear that wrist-worn optical HR monitors are highly susceptible to motion artifacts from physical activity. In addition, the optical HR monitors' typical battery life of about one day for continuous monitoring is insufficient for military operations that could last for three days or more. Because of these two key deficiencies, we are now investigating how to improve robustness to motion artifacts while reducing the power of optical HR monitors.

About the Authors

Brian A. Telfer is a senior staff member in the Human Health and Performance Systems Group at Lincoln Laboratory. He joined the Laboratory in 1995 after working at the Naval Surface Warfare Center, where he was funded with an Office of Naval Research Young Scientist Award. His contributions to the Laboratory's R&D have been in signal processing, machine learning, artificial intelligence, and systems analysis, initially for ballistic missile defense and for the past seven years for bioengineering, particularly for physiological status monitoring. The focus of this work has been on prototyping and technology transition, with successful transitions leading to government contracts to industry for several hundred million dollars and smaller industrial contracts that have resulted in several hundred (and growing) physiological monitoring units. He has organized body sensor network workshops and has served as technical program cochair for the International Body Sensor Networks Conference in 2013 and 2015. He is a senior member of the IEEE and a member of the IEEE Technical Committee on Wearable Devices. He has also served on several Lincoln Laboratory internal funding selection boards and as the technical lead for the Laboratory's Introduction to Radar Systems course and other radar courses. He has led several studies for senior leaders in the U.S. government and has coauthored more than 70 publications. He holds a bachelor's degree in electrical engineering from Virginia Polytechnic Institute and State University and master's and doctoral degrees in electrical engineering from Carnegie Mellon University.

Kate Byrd is an associate technical staff member in the Advanced Sensors and Techniques Group at Lincoln Laboratory. Her work focuses on small RF systems and has included using optimization methods to design miniature wideband antennas for communication systems and designing small radars to detect the movement of disaster survivors under rubble. She originally joined the Laboratory in 2015 and worked in the Bioengineering Systems and Technologies Group. She was involved in biosensor design, testing, and data analysis. This work resulted in a patent for hardware designed to collect noise exposure and an R&D 100 Award for a system that monitors and prevents musculoskeletal injuries in soldiers. In 2018, she was accepted to the Lincoln Scholars tuition grant program for graduate studies. She holds a bachelor's degree in bioengineering from Clemson University and a master's degree in electrical engineering from Harvard University.

Paula P. Collins is the assistant leader for Lincoln Laboratory's Human Health and Performance Systems Group. She oversees efforts to advance integrated wearable sensing for human physiological, cognitive, and psychological performance monitoring, primarily in support of military operations. Prior to this position, she was a member of the technical staff, leading numerous projects to develop and evaluate a broad range of advanced technologies for biomedical applications; intelligence, surveillance, and reconnaissance (ISR) architectures; and air vehicle survivability. Over her career at the Laboratory, she has gained extensive experience with multimodal systems-of-systems sensor architectures, complex large-scale human-in-the-loop data collection, and data analysis for objective evaluation of proposed advanced capabilities. She holds a doctorate in experimental particle physics from the University of Pittsburgh.

Understanding Noise-Induced Auditory Damage

Christopher J. Smalt, Paul T. Calamia, Shakti K. Davis, Joseph J. Lacirignola, and

Thomas F. Quatieri

Noise-induced hearing injuries are pervasive in the United States, and large numbers of military personnel who train or carry out missions under high-noise conditions have particularly been affected. Understanding the impacts of noise on the auditory system and developing metrics to predict the likelihood and severity of these impacts are key to developing hearing protection devices that will prevent, or mitigate, hearing impairments. Lincoln Laboratory is performing R&D to improve this understanding and to create computational models that can inform the development of effective hearing protection for warfighters.

» **Noise-induced hearing injuries** (**NIHI**), including tinnitus and hearing loss, are among the most common disabilities suffered by active-duty warfighters and reported by military veterans. Although noise-exposure risks can vary greatly across military branches and occupations, most service members are regularly exposed to noise hazards through training and operational duties during their careers, and the resulting auditory impairments are unfortunately common. For example, tinnitus and hearing loss have long been, and continue to be, the two most common disabilities compensated by the Department of Veterans Affairs (VA). In fiscal year 2016 alone, the VA awarded 149,429 new tinnitus compensation benefits and 77,622 new hearing loss disability benefits [1].

Hearing impairment and auditory-perceptual dysfunction resulting from noise exposure can reduce situational awareness by degrading sound detection thresholds, localization accuracy, and speech intelligibility. For the warfighter, these impairments not only can threaten mission success and survival, but also can adversely affect post-service life. Ironically, hearing protectors can induce similar detrimental effects through the attenuation that they provide, leading to their lack of acceptance by warfighters. A small number of studies have investigated the effects of hearing impairment or hearing protection on military operational performance during training and simulated operations, and in all cases, these studies found not only reductions in the subjects' ability to communicate, detect, and localize sounds [2] but also lethality in a combat exercise [3].

Such effects may result directly from peripheral auditory impairments that reduce input to the brain, indirectly from central deficits that compromise the neural processing of auditory information, or from a combination of both.

To limit noise exposure and reduce the risk of NIHI, hearing-protection devices (HPDs) are made available to warfighters in training exercises and during operational duties. However, these devices often are disliked and consequently are not worn, in part because the devices can compromise auditory situational awareness and degrade speech intelligibility [4]. Another challenge in addressing the risk of NIHI is the difficulty in accurately quantifying the noise exposure experienced by a warfighter during a mission or training exercise, or even in a standard work day. Noise-dosimetry devices are intended to measure noise exposure levels, but a variety of confounding factors—including device placement, environmental noise conditions, suboptimal hardware components, and inadequate predictions of auditory risk—can result in the dosimeters producing a poorly informed assessment of NIHI risk for a given situation.

Auditory Physiology and Damage Mechanisms

While high sound-pressure-level events such as explosions are capable of causing damage to the outer or middle ear, e.g., a ruptured tympanic membrane, i.e., ear drum, NIHIs are most often considered with respect to the inner ear, or cochlea. Mechanical signals (transmitted from sound-pressure wave energy in the middle ear) are transduced into electrochemical signals within the cochlea, and these signals are communicated, in turn, as nerve impulses via the auditory nerve to the brain (Figure 1). On the basilar membrane within the cochlea is the organ of Corti, lined by specialized primary auditory receptor inner and outer "hair" cells. Hair cells play a crucial role in hearing: in response to mechanical vibration that displaces their hair-like stereocilia, they transduce that energy into electrical current to the auditory nerve. Noise-induced hearing loss can occur when strong pressure waves (from very loud sounds) damage or destroy hair cells and their transduction machinery, or their associated neural connections. In humans, damaged and lost hair cells cannot regenerate, so the resulting hearing impairment is permanent.

Historically, hair-cell loss has been the focus of most research efforts to explain noise-induced hearing loss. Inner hair cells transform sound waves into signals for relay to the auditory nerve, brainstem, inferior colliculus, thalamus, and auditory cortex. Each inner hair cell is innervated by many afferent (ascending) nerve fibers—approximately 30,000 per ear—which are activated by neurotransmitters released across the synaptic space. Stimulation of the inner hair cells is the primary trigger for acoustic information to be sent to the brain. Inner hair cells can be damaged directly by loud noise and have additional susceptibility for noise-induced damage at their synapses with the auditory nerve.

Outer hair cells function primarily to amplify sound in a frequency-dependent, compressive, nonlinear
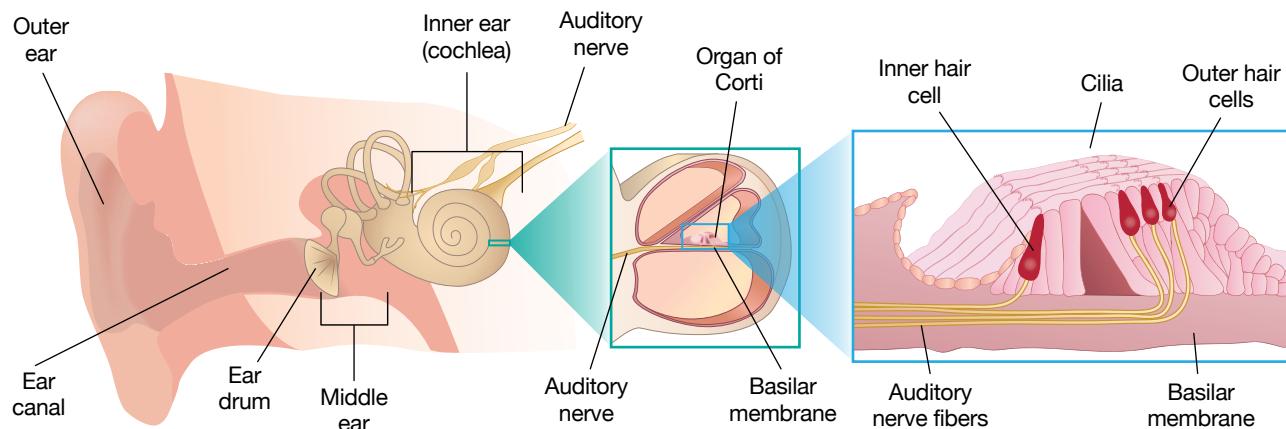
FIGURE 1. The auditory periphery includes the outer, middle, and inner ears (left). At center is an exploded diagram of the cochlea, and at right is a cross-section of the organ of Corti within the cochlea (reused from [5]).

manner such that a wide range of sound pressures are encoded by small displacements of the inner hair cells. In this way, outer hair cells "fine-tune" frequency selectivity along the basilar membrane. Outer hair cells also receive efferent (descending) inhibitory innervation from the brain stem; this innervation attenuates and reduces the frequency selectivity of afferent signals. The interplay of afferent and efferent innervation acts as a feedback loop, allowing the brain to adjust hearing sensitivity on the basis of frequency and sound level. Outer hair cells are thus critical to the reception of quiet sounds, sound frequency discrimination, and perception of sound in noisy environments. Loss of outer hair cells leads to what is most commonly thought of as the noise-related damage that affects general "hearing ability," but can also manifest as hearing impairment beyond changes in audibility, e.g., reduced speech comprehension.

Hearing ability typically is assessed as changes in audiometric thresholds at various frequencies in clinical tests known as audiograms, which are indicative of hair-cell damage or loss. When hearing function tests appear normal, the conclusion is often made that no significant or permanent injury has occurred. However, recent research in animal models suggests that well before noise exposure causes hair-cell death, it can cause extensive injury to auditory nerve synapses that is not reflected in any appreciable audiogram changes. In animal experiments, researchers at the Massachusetts Eye and Ear Infirmary have shown that dramatic cochlear synaptopathy (loss of synaptic connections) can occur after exposure to continuous noise [6–9]. These studies indicate that as many as half of the auditory nerve fibers are lost after noise exposure at levels that cause neither permanent damage to hair cells nor permanent threshold elevations as measured by an audiogram. The initial synaptic loss can be observed within hours of noise exposure while subsequent death of the spiral ganglion (auditory nerve) cell bodies continues slowly over months or years. Despite histological degeneration of the auditory nerve, sensory function (as measured by an audiogram) remains normal. However, transfer of auditory information to the brain may be compromised, resulting in diminished perceptual processing. In particular, synaptopathy is thought to be associated with difficulty in understanding speech in noisy backgrounds. This difficulty occurs because the type of synapses that

are most susceptible to damage encode suprathreshold sound information. Since an audiogram measures hearing thresholds, i.e., the softest sounds a person can hear, it cannot detect synaptic loss. It is not currently possible to noninvasively observe synaptic damage in living human subjects, so the prevalence and severity of this form of auditory damage in the population remain uncertain.

Noise Exposure

Hazardous noises arise from a wide variety of sources. In the military, some common sources of noise include aircraft, land vehicles, and naval vessels, as well as weapons fire and other explosive blasts. To understand noise characteristics and their potential effects on auditory health, it is useful to classify noise into two general categories as illustrated in Figure 2: continuous and impulsive. Continuous noise is relatively uniform over time, exhibiting only minor fluctuations in level or frequency content. Engine rooms and aircraft cockpits are examples of military work environments in which loud continuous noise is a concern. Impulse noise is characterized by a sharp burst of acoustic energy with rapid rise and decay times. Weapons fire and other explosions are examples of impulse noise. Heterogeneous combinations of both categories are known as complex noise.

Beyond categorizing the type of noise, alternative approaches have been used to quantify the noise to which an individual is exposed and to set guidelines for safe exposure. To those ends, noise dosimetry involves measuring sound-pressure levels in an environment to estimate an individual's exposure throughout a day, work shift, or event of interest. Noise dose typically is estimated in terms of acoustic energy in conjunction with the equal-energy hypothesis (EEH), which assumes that accumulated noise energy is sufficient to determine risk of NIHI without consideration of the underlying temporal or spectral characteristics. Under the EEH, two exposures are equivalent if the respective average noise levels and durations comply with a specified exchange rate. For example, a 3-decibel (dB) exchange rate often is employed such that a halving or doubling of the exposure time is accommodated with a +3 or −3 dB adjustment, respectively, to the allowable noise level.

To conserve individuals' hearing in industrial and military settings, regulatory agencies, such as the National Institute for Occupational Safety and Health,
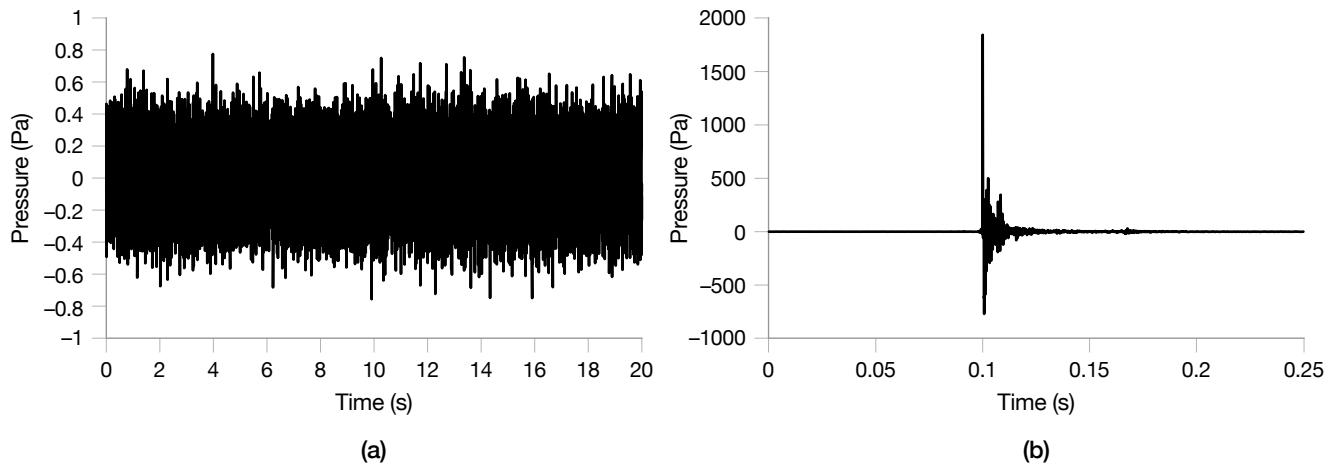
FIGURE 2. The waveforms depicted illustrate the relatively uniform level of continuous noise (a) and the short rise time, high peak level, and rapid decay of impulsive noise (b).

and military branches under the Department of Defense (DoD) Hearing Conservation Program have recommended guidelines on the maximum allowable daily noise exposure. For example, the current military standard sets a limit of 85 dBA[1] for a duration of eight hours for continuous noise exposure, where the exposure duration and level may be traded off to satisfy an equal-energy criterion using a 3 dB exchange rate.

Noise-exposure measurements can be compared against established limits or other criteria to determine a need for hearing protection or to predict the risk of hearing loss. Numerous damage-risk metrics have been proposed to quantify harmful aspects of noise exposure, but risk-assessment metrics remain an active area of research because no single metric is considered adequate across the spectrum of noise conditions. The most common damage-risk metric is a time-weighted average of the A-weighted noise level:

$$L_{\text{Aeq,T}} = 10\log_{10}\left[\frac{1}{T}\int_T \frac{p_A^2(t)}{p_0^2}dt\right]$$

where T represents the exposure duration, p_0 = 20 μPa is the reference pressure level, and $p_A(t)$ is the A-weighted pressure-time waveform. As noted previously, the

military noise exposure limit is 85 dBA for an eight-hour period, that is, $L_{\text{Aeq,8h}}$ R 85 dBA [10].

While $L_{\text{Aeq,8h}}$ has wide acceptance as a damage risk metric for continuous-noise exposure, many concerns have been raised that it is not adequate for predicting hearing damage from complex or impulsive noise [11]. One concern is that $L_{\text{Aeq,8h}}$ and other energy-based metrics ignore much of the temporal and spectral structure of the noise; yet, evidence suggests that some of these features influence the damage severity from impulsive and complex noise. Hamernik et al. [12] showed that $L_{\text{Aeq,T}}$ under-predicts hearing damage when continuous and impulsive noise are combined, and other studies suggest that impulsive exposures with predominantly low-frequency energy may be less hazardous than an equal-energy impulse dominated by higher frequencies [13]. Furthermore, the linear relationship between energy and permanent auditory threshold shifts only holds for noise levels up to about 140 dB [14]. Above this level, nonlinear operations may be necessary to translate the energy metric into auditory damage. In response to these concerns, several complementary or alternative metrics have been proposed for impulsive noise.

Recently, the military adopted the Auditory Hazard Assessment Algorithm for Humans (AHAAH) as part of the noise-limits design standard, MIL-STD-1474E [15]. The AHAAH electro-acoustic model developed by the U.S. Army takes an impulsive noise waveform as input and calculates an output value in auditory risk units (ARUs) that represents energy reaching the inner ear

---

[1]The A in dBA indicates that a filtering process meant to mimic the varying sensitivity of the human auditory system with respect to frequency, known as A-weighting, is used in the acoustic energy calculation.

CHRISTOPHER J. SMALT, PAUL T. CALAMIA, SHAKTI K. DAVIS, JOSEPH J. LACIRIGNOLA, AND THOMAS F. QUATIERI

[16, 17]. While AHAAH has been adopted in the standard for military acquisitions, the military acknowledges some limitations of this model, including the need for further validation. Zagadou et al. [18] recently suggested a number of modifications to the AHAAH parameters and hazard assessment calculation that yield a better fit to existing human blast exposure data. Another area of active research on AHAAH is evaluating the assumption that the middle-ear muscle contraction can provide significant protection against acoustic damage [19]. The so-called "warned" condition of the AHAAH assumes that this acoustic reflex is engaged prior to exposure and that the associated muscle contraction reduces energy transfer to the inner ear by as much as 20 dB, compared to the "unwarned" condition [20].

A limitation with the aforementioned metrics is that while $L_{\mathrm{Aeq,8h}}$ is intended for assessing continuous noise and AHAAH is for impulsive noise, no guidelines exist for combining the predicted risk of complex noise in which both continuous and impulsive noise pose a hazard. Several recent studies have sought to model auditory damage from complex noise exposures that may be more realistic to military and industrial settings. One concept to address the inaccuracy of the equal-energy hypothesis is a kurtosis correction factor for time-weighted average noise levels [21–23]; compared to the uncorrected, this concept has been shown to improve correlation against permanent auditory threshold shifts in chinchillas. Recently, Sun et al. [24] proposed an alternative kurtosis-based energy metric that adaptively elevates the effective energy in impulsive noise environments and reverts to the conventional A-weighted calculation in continuous noise environments. This approach is promising because it aims for a unified metric that appropriately adapts to the noise environment, but further study is needed to validate kurtosis-corrected energy metrics over more datasets, including those from complex military noise environments.

Noise Dosimetry for Military Environments

Noise dosimetry involves the collection of environmental noise data and the calculation of exposure levels with free-field, on-body, and/or in-ear devices. Free-field noise surveys typically characterize the noise levels of an environment, but accurately translating such a survey to the dose for an individual can be challenging. For example, sound-pressure levels at the ear drum can differ dramatically from those in an arbitrary free-field location (by 10 dB or more), depending on the exact positioning of body and ear relative to the noise sources [25]. Many modern, small-form-factor commercial off-the-shelf (COTS) dosimeters can be worn on the body (preferably in close proximity to the ear) to directly measure the dose in the vicinity of an individual, but they typically lack the dynamic and frequency ranges necessary for military use. Another potential measurement location for noise is in the ear canal. This measurement can be achieved by integrating a microphone into a hearing-protection device to allow for characterization of the noise exposure when hearing protection is worn. This configuration is referred to as an in-ear measurement, in contrast with the on-body measurement that may follow the individual but does not account for the noise attenuation of the hearing-protection device.

Hardware requirements for a dosimetry device vary for different noise types and environments. Military noise environments are complex, with both continuous and impulsive noise. The latter, for example from weapons fire, typically is the most demanding with respect to dosimeter design because of its highly dynamic nature and extreme levels. This challenging set of characteristics drives the need for a broadband dosimetry device with a high sampling rate and a wide dynamic range to avoid clipping or distortion from large blasts. Typical commercial noise dosimeters operate up to 140 dB peak sound-pressure level (SPL) and cover a frequency range similar to that of human hearing [26]. However, weapons fire, blasts, and other impact noises can exceed this SPL limit, and impulses can exhibit acoustic bandwidths extending well beyond the audio spectrum because of their short durations [27, 28].

Size, weight, and power are important considerations in designing a dosimetry device that is wearable and capable of measuring noise for a significant length of time (more than eight hours). For a small package suitable for an on-body or in-ear system, the trade-off typically will be between recording fidelity (driven by sample rate and bit depth) and recording duration (driven by battery life and memory capacity). While many commercial devices are available in a wearable form factor, they are designed for industrial noise environments and do not have adequate microphones or sampling rates to characterize impulsive

noise with peak levels above 140 dB. Portable commercial audio recorders are one alternative that can be used to capture high-fidelity noise exposures with external microphones capable of measuring high-SPL noise; however, such devices are often bulky and have many settings and stressing power requirements to support the high-SPL external microphones. Furthermore, audio recorders do not support onboard processing to calculate noise metrics in real time. Smartphones have been successfully used to measure some noise environments [29] but are limited by a low sample rate and dynamic range. External microphones can be paired with a smartphone to increase the maximum SPL, but these require separate power and circuitry.

A comparison of some important characteristics for dosimetry devices is shown in Table 1. The COTS dosimeter column represents a state-of-the-art commercial dosimeter designed for industrial noise environments. The limitations on peak SPL and sample rate motivate alternative setups for exposure measurements in military environments. Two alternatives are explored in the middle columns and are described later in more detail. The final column represents a notional ideal device based, in part, on the impulse noise measurement requirements from the military specifications in MIL-STD-1474E. The ideal peak SPL of 175 dB is guided by the peak levels expected across a variety of military ordnance.

Lincoln Laboratory Dosimeter Prototyping
In response to the gap in available COTS devices to support noise measurements in military environments, Lincoln Laboratory began designing and prototyping a noise dosimeter in 2013 to support an operational noise collection on U.S. Marines in Afghanistan. The study was fielded by the Marine Expeditionary Rifle Squad (MERS) as part of a joint protocol with the U.S. Army Research Institute of Environmental Medicine.

Table 1. Performance Characteristics for Various Noise-Dosimetry Hardware Options

METRIC	COMMERCIAL OFF-THE-SHELF (COTS) DOSIMETER	AUGMENTED COTS RECORDER	LINCOLN LABORATORY SECOND-GENERATION DOSIMETER PROTOTYPE	IDEAL DEVICE FOR MILITARY DOSIMETRY
Peak sound-pressure level	140 dB	172 dB with external microphones	175 dB	175+ dB
Sample rate	20 kHz	96 kHz	128 kHz	192 kHz
Form factor	Clip to clothing	Small pouch	Small pouch	Clip to clothing
Maximum collection time	24+ hours	7 hours	8 hours	8+ hours
Measure in-ear and on-body noise	On-body only	Both with custom microphone housings	Both	Both
On-board processing	Yes, but limited metrics	No	Yes	Real-time calculations of noise exposure metrics

■ Excellent ■ Acceptable ■ Low-quality solution

The Laboratory's first-generation prototype had a primary objective of supporting the collection of data from small-arms fire through increasing the maximum SPL and sample rate. While this prototype was an improvement over the existing COTS dosimeters, the initial prototype fell short on some of the ideal requirements listed in Table 1. Building on the lessons learned from the first-generation system, Laboratory researchers developed a second-generation noise dosimeter designed to be more portable and capable of satisfying the instrumentation specifications of MIL-STD-1474E. Both generations of prototypes are shown in Figure 3. The second-generation prototype was funded jointly by MERS and the U.S. Army Combat Capabilities Development Command Soldier Center (formerly known as the U.S. Army Natick Soldier Research, Development, and Engineering Center). Several auxiliary sensors were integrated into the second-generation device to add a capability for simultaneously measuring other environmental data, including location (GPS), temperature, barometric pressure, and acceleration. The second design also included onboard data processing via a Xylinx Zynq, which has a field-programmable gate array and a dual-core ARM processor, as well as co-located accelerometers at each microphone to help detect and screen out microphone artifacts in a post-processing step.

The measurement quality of the second-generation dosimeter was verified by directly comparing the dosimeter to a reference laboratory-grade data-acquisition system (National Instruments, 24-bit, 200 kHz sample rate) for a series of high-SPL impulse-noise events. Figure 4 shows the test configuration and an example measurement that illustrates the close agreement between the prototype and the reference system for a 161 dB impulse generated from a compressed-air shock tube.

The peak SPL, $L_{eqA,100ms}$ and median difference in the 1/3-octave-band levels of the prototype were within ±1.5 dB of the reference system for eight blast measurements with peak SPLs in the 160–179 dB range.

Augmented COTS Audio Recorder

During the development of the second-generation dosimeter prototype, several field collection opportunities arose that required an interim solution for measuring noise during military exercises. An augmented COTS recorder was constructed using a two-channel, 24-bit TASCAM DR100-MKIII recorder with a sampling rate of 96 kHz (Figure 5). With the addition of dual microphones and custom housings to support in-ear and on-body noise measurements, this device supports a peak SPL of 172 dB and approaches several of the ideal requirements of Table 1. The device is primarily limited by a lack of onboard processing and a bulky form factor. Laboratory tests were performed to validate the measurement quality through a comparison with the reference laboratory-grade National Instruments data-acquisition system.

To enable in-ear measurements with one of the COTS microphones connected to this augmented recorder, Lincoln Laboratory designed a custom, 3D-printed

Electronics housing
1.8 in
3 in
0.9 in

Microphone
1 in

(a)

Electronics housing
3 in
5 in
1 in

Microphone
0.5 in

(b)
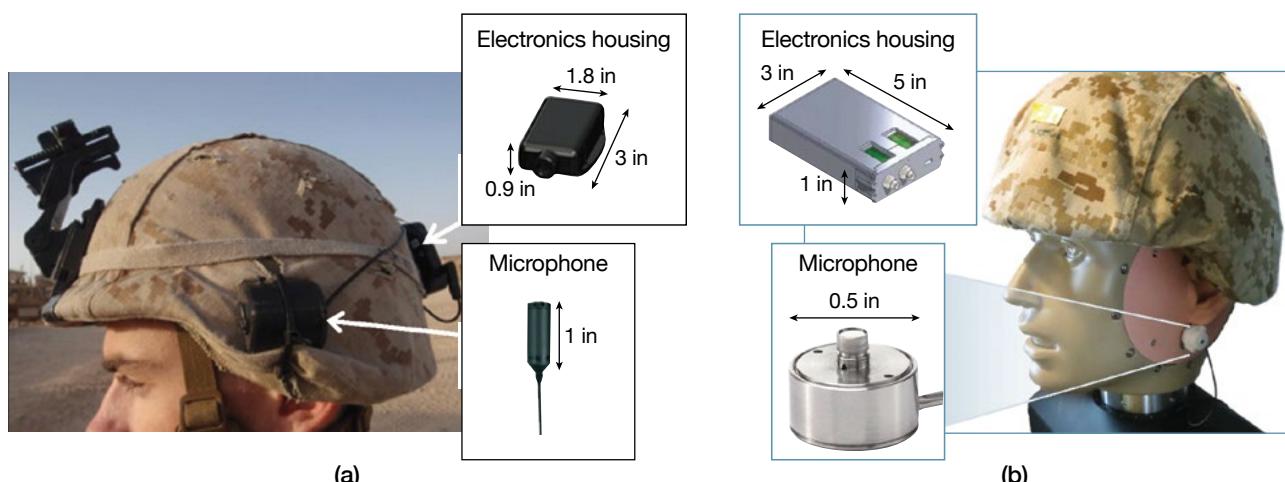
FIGURE 3. The first- (a) and second-generation (b) Lincoln Laboratory noise dosimeters are depicted.

Peak SPL: 161.7 dB

— NI PXI 4462   — Lincoln Laboratory's
second-generation dosimeter

(a)   (b)
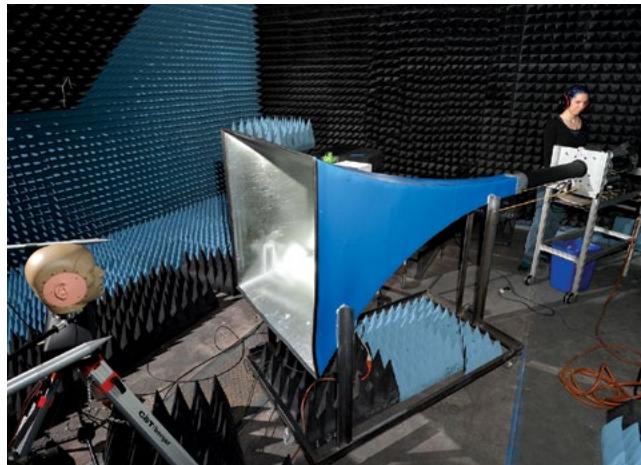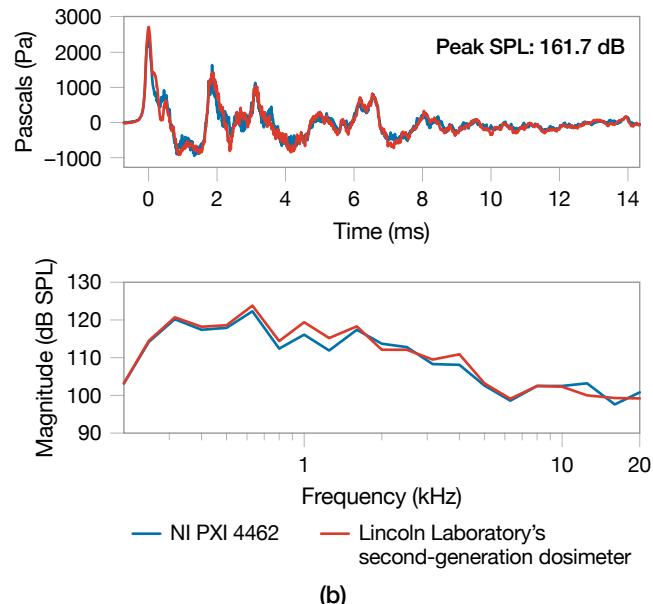
FIGURE 4. The shock tube measurement setup is seen in (a), and an example comparison of Lincoln Laboratory's second-generation dosimeter to the reference data-acquisition system measurements is depicted in (b). Note that the two systems have comparable performance in terms of peak amplitude and overall blast waveform shape (upper b) and blast spectral content (lower b).

housing that couples with a military-grade hearing protector ear-tip (Figure 6). Prior to test subjects' using this device as hearing protection, the impulse peak insertion loss was measured in accordance with the ANSI HPD testing standard [30], repeating the same shock tube and acoustic test fixture setup illustrated in Figure 4a.

The blast tests indicated that the in-ear microphones integrated with the foam ear-tip HPDs achieved an impulse peak insertion loss 45 dB or higher for blasts in the range of 148–172 dB. This level of suppression provides sufficient protection for many military exercises, including rifle training. Furthermore, the measurement quality of impulse noise captured by the augmented COTS recorder is very similar to the measurement quality of the reference laboratory-grade data acquisition system. Over a series of 18 blasts, a root-mean-square (RMS) error of 1.2 dB was found between the peaks measured by the augmented COTS recorder and those measured by the reference system.

One opportunity to collect and characterize military noise exposure occurred during Marine rifle training in which instructors are subjected daily to thousands of rounds of rifle fire. The augmented COTS recorder with both on-body and in-ear microphones was used to collect noise exposure measurements for seven rifle

FIGURE 5. As an alternative to the second-generation dosimeter, Lincoln Laboratory augmented a commercial noise recorder (above) with dual microphones and custom housings so that it could be used during a military exercise.

instructors over a two-day period at Marine Corps Base Camp Pendleton.

Representative results of the daily noise exposure for two rifle range instructors are shown in Table 2. Several noise-exposure metrics were calculated on the basis of data from both the in-ear and on-body microphones. Subjects were asked to wear the in-ear microphone in place of their normal hearing protection during the

CHRISTOPHER J. SMALT, PAUL T. CALAMIA, SHAKTI K. DAVIS, JOSEPH J. LACIRIGNOLA, AND THOMAS F. QUATIERI
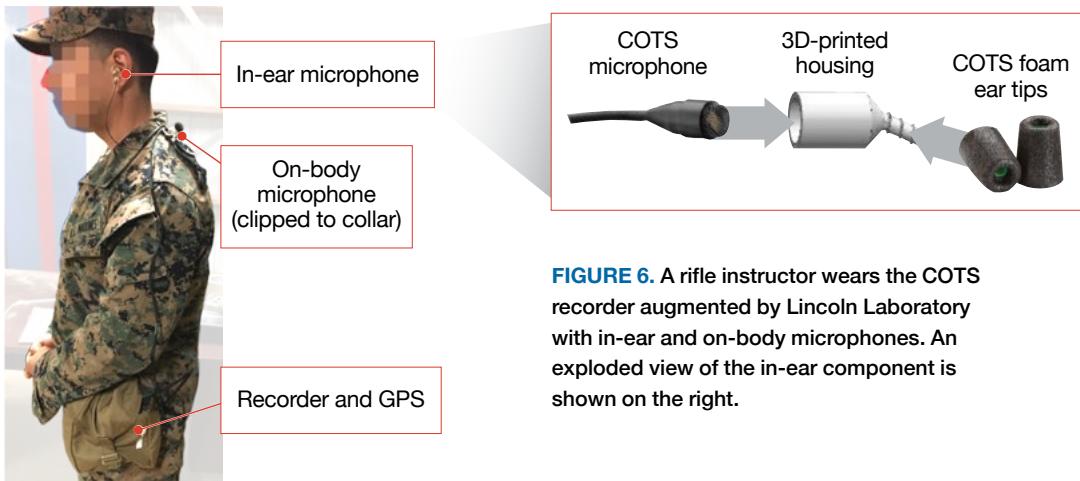
FIGURE 6. A rifle instructor wears the COTS recorder augmented by Lincoln Laboratory with in-ear and on-body microphones. An exploded view of the in-ear component is shown on the right.

collection. The on-body microphone measurements represent the potential noise exposure that individuals would have experienced had they not worn hearing protection; the in-ear measurements represent the actual noise that reached the ear canal with the HPD in place. Eight-hour equivalent energy and the corresponding dosage were computed along with the AHAAH ARU under the unwarned assumption. Without hearing protection, each instructor would be exposed to more than 1,000 impulses per day, each with peak SPL above 140 dB. The eight-hour equivalent noise energy on the rifle range without HPD is more than 15 dB above the DoD limit, corresponding to exposure dosages that are 30 to 45 times higher than the DoD limit. With hearing protection, the exposure metrics are considerably reduced. However, variability in the individual fit of the HPD can lead to dramatic differences in protection.

Of the two instructors represented in Table 2, Subject 5 achieved better overall suppression of the noise energy because of a better fit of the HPD, and he received a dose well below the limit. However, at a few points during the day, he briefly removed the HPD and, consequently, was exposed to a few rifle shots without hearing protection. This handful of unprotected impulses was enough to raise his AHAAH ARU count above the daily limit, despite his overall energy dosage being within the limit. In contrast, Subject 4 did not attain an equally good HPD fit, and as a result, the accumulated acoustic energy that reached his ear canal was more than twice the daily limit. However, this subject wore his HPD throughout the day, avoiding any unprotected impulse

exposures. This consistent protection throughout the day held the AHAAH ARU at zero, despite the overall energy exceedance.

The on-body noise exposure measurements from the Marine rifle instructors indicated high-risk noise conditions for both the continuous and impulsive metrics. Wearing hearing protection reduced the exposure risk, but the level of protection can vary dramatically, depending on how well the HPD was inserted in the ear canal or whether the HPD was removed, even briefly, during the day.

Future Noise Dosimetry Considerations
Capturing both in-ear and on-body noise measurements is important for developing relevant noise-exposure models. In addition, collecting coordinated audiometric tests on warfighters during military operations or training could generate important datasets for evaluating existing noise metrics and validating new ones. To date, most military noise-exposure standards are typically validated on animal tests of blast overpressure exposures or a single human study of blast overpressure exposures conducted on more than 200 military volunteers in the early 1990s [31]. Additional dosimetry and audiometric collections during military training or operations could provide valuable data to help validate exposure metrics and standards over a wide variety of military noise conditions. In addition to improving noise-exposure standards, future data collections of this type may help to inform individual susceptibility for NIHI by including other physiological and genetic factors.

Table 2. Representative Metrics for Two Rifle-Range Instructors at Camp Pendleton

NOISE EXPOSURE METRIC	WITHOUT HEARING-PROTECTION DEVICES		WITH HEARING-PROTECTION DEVICES		DEPARTMENT OF DEFENSE LIMITS
	SUBJECT 5	SUBJECT 4	SUBJECT 5	SUBJECT 4	
$L_{Aeq,8h}$	100 dBA	101 dBA	80 dBA	88 dBA	≤85 dBA
Dose	2995%	4479%	33%	219%	≤100%
Auditory Hazard Assessment Algorithm for Humans auditory risk units (ARU) (unwarned)	152,129	161,291	1004	0	≤200 ARU
Number of impulses ≥140 dB	1145	1393	6	0	n/a

■ DoD noise limits were exceeded   ■ Within DoD noise limits

Hearing Protection

Warfighters often do not use hearing protection provided to them for a variety of reasons—comfort, poor integration with other headgear, and compromised auditory situational awareness. Many of these concerns are echoed by participants in noisy, nonmilitary jobs [32], so a deeper understanding of HPD performance and limitations, with the ultimate goal of improving their usability and efficacy, is relevant to a wide swath of the population.

Studies of the effects of hearing protection on auditory situational awareness typically include some combination of tasks related to sound-localization performance, detection thresholds, and speech intelligibility [33, 34]. We are focusing on the first of these, along with a novel assessment of the additional cognitive load (listening effort) induced by wearing an HPD.

Devices and Subjects

Hearing-protection devices fall into one of two main categories: passive or active. Passive HPDs rely on occlusion of the ear canal, along with acoustic absorption and/or impedance mismatch, to prevent sound from impinging on the eardrum. The form factor may be over

the ear, e.g., an earmuff, or in the ear, e.g., an ear plug. Active HPDs also rely on occlusion to limit the acoustic energy entering the auditory system, but in addition they contain an outward-facing microphone and an inward-facing loudspeaker, along with relevant signal processing algorithms, to enhance their performance. The most basic mode of an active HPD involves attenuating the signal at the microphone and passing it through to the loudspeaker otherwise unchanged. More sophisticated devices provide features such as nonlinear dynamic-range compression, active noise cancellation, and modes designed to address different noise types (e.g., impulse noise or continuous noise). Our HPD evaluation included two passive and three active devices, with the open ear as a reference.

Data-Collection Platform

Data were collected in a double-walled sound-isolating booth in which 24 loudspeakers were positioned in an eight-foot-diameter ring approximately five feet above the floor, as shown in Figure 7. The speakers were evenly spaced and pointed toward the central listening position; they were covered by an acoustically transparent cloth so their locations cannot be determined visually. Additionally,
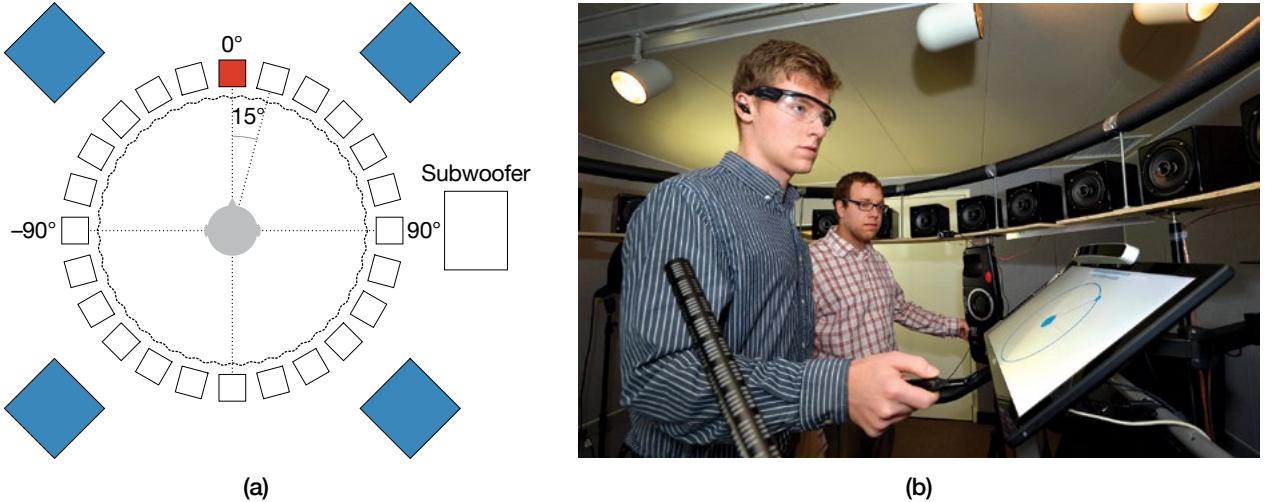
FIGURE 7. The experiment setup (a) includes 24 stimulus speakers evenly spaced in a ring with a diameter of eight feet. Larger speakers for background-noise output are shown in blue outside the ring, with a subwoofer shown on the right. The dashed line represents an acoustically transparent screen. The red-shaded speaker marks the direction directly in front of the listener, i.e., 0 degrees azimuth. The photograph in (b) shows the actual system, including the touch screen, with the acoustically transparent cloth removed to expose the loudspeakers.

four large speakers were placed in the corners of the booth to generate background noise if desired. Each test subject was seated in the center of the ring, positioned such that his/her ears were approximately at speaker level, with a 20-inch touchscreen positioned to display instructions and capture the subject's responses to tasks.

Sound Localization

The target stimulus used for the localization task was a recording of an AK-47 rifle cocking, an acoustic signal of approximately one second in duration with a broad spectrum that has been used in previous HPD localization studies [34]. The stimulus playback was calibrated at the listening position so that its level was matched across all 24 speakers and presented at 65 dBA SPL. Three background noise conditions were used for the sound localization task: quiet (<20 dBA ambient noise) and a Blackhawk helicopter noise calibrated at 60 dBA and at 80 dBA. These noise levels were chosen to span the range from "clearly audible" to "very difficult to localize" for the stimulus. The stimulus was played from one speaker at a time in random order, and each speaker was used twice for a total of 48 trials per background-noise level. The subject used a positional tracking device to point in the direction from which she/he heard the sound on each trial. The subject was shown the direction that she/he

was pointing on a computer screen in real time to prevent errors in the pointing direction caused by any obstruction of the sensor by the body. Localization performance was assessed by measuring the mean angle error (MAE) between the target speaker and the subject's response in degrees azimuth. Quadrant errors, which manifest as front/back or left/right confusions, also were quantified.

Figure 8 shows the MAE in localization for the open ear and five hearing-protection devices across all subjects for each of the three background-noise conditions. The mean error for each subject was computed across the 48 trials per noise condition and HPD. In addition to the degradation of localization accuracy, quadrant errors also were evaluated as shown in Figure 9. The space around the listener is divided into four quadrants defined as: right-front (0° R azim. < 90°); right-rear (90° R azim. < 180°); left-rear (180° R azim. < 270°); left-front (270° R azim. < 360°). Quadrant errors occur when a subject's localization estimate is in a different quadrant from the actual stimulus loudspeaker (typically a front/back error rather than a left/right error), and the difference between the estimated and actual azimuth is greater than 30 degrees.[2] A

---

[2]The 30-degree minimum constraint was imposed to differentiate actual quadrant errors from localization blur near the quadrant boundaries.

two-way analysis of variance (ANOVA) performed on the mean localization MAE revealed that there was an effect of HPD ($F_{(4, 47)} = 45.0$, $p < 0.001$), noise level ($F_{(2, 23)} = 104.5$, $p < 0.001$), and a significant interaction ($F_{(8, 95)} = 2.189$, $p < 0.035$) between the two. Once quadrant errors were removed, the interaction effect was no longer significant ($F_{(8, 95)} = 0.315$, $p = 0.95$). A comparison of Figure 8 and Figure 9 suggests that the dominant cause of degraded localization performance with an HPD is the prevalence of quadrant errors. The implications of localization errors in terms of warfighter performance are an active area of research.

Cognitive Load

A dual-task paradigm was employed to evaluate the effect of the HPDs on cognitive load. The primary task involved the evaluation of speech intelligibility in 75 dBA noise by using the modified rhyme test (MRT), a six-alternative forced-choice task for which each trial consists of the carrier phrase "Please select the word" followed by a target word [35]. The target word, played from the speaker directly in front of the subject, rhymed (i.e., either the first or last consonant varied) with a number of word selections presented on the touch-screen response pad. The subject had three seconds to select the word he/she heard before the next trial began. The secondary task required the subjects to remember the five previously selected words in the MRT. In addition, visual reaction times in response to the lighting of a red LED light above the touch screen were measured throughout the experiment. Visual reaction time has previously been shown to be sensitive to mental effort [36] and has been established as a metric for predicting cognitive load.

To control for variations in individual reaction time and alertness over the multiple days of testing, reaction times were acquired without the MRT or word-recall task. Noise and stimulus levels were kept in the same configuration as during the dual-task testing, but subjects were instructed to ignore the speech and noise, and only respond to the visual stimulus as quickly as possible. Dual-task MRT blocks were alternated with reaction-time-only blocks in an A-B-A-B fashion. The median reaction time over both reaction-time-only blocks was taken as the single-task visual reaction-time baseline for each hearing-protector condition. Reaction time and word recall were initiated at random intervals (every 7
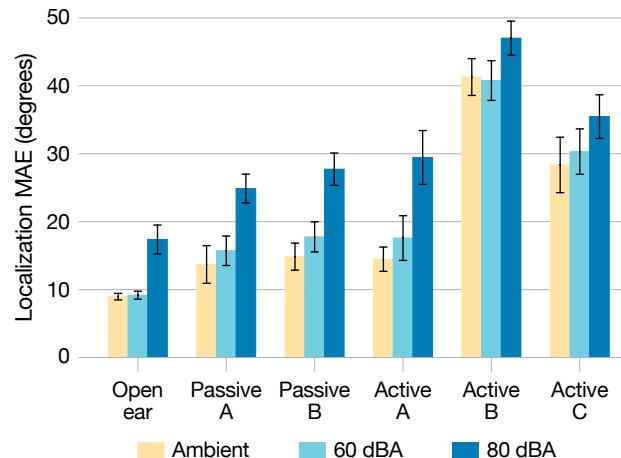
FIGURE 8. The localization performance in terms of mean angle error (MAE) is given for the three noise conditions (ambient, or quiet; helicopter noise at 60 dBA and at 80 dBA) for the open-ear condition and for five hearing protection devices (two passive HPDs and three active HPDs). Performance with all devices degrades with increasing background noise. The large errors induced by the Active B and C devices, caused mainly by quadrant errors as shown in Figure 9 below, are particularly detrimental to auditory situational awareness.
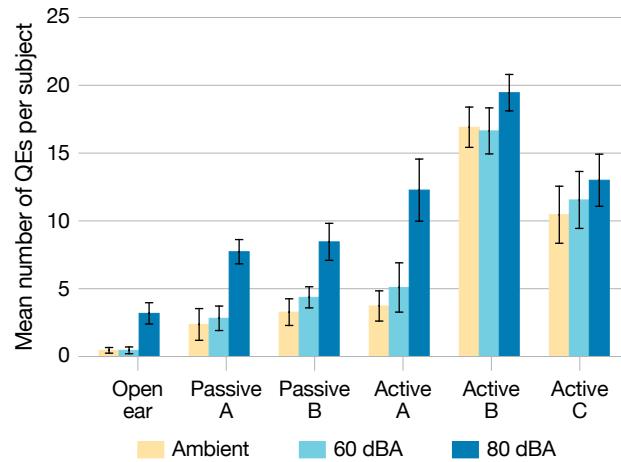
FIGURE 9. Localization performance in terms of quadrant errors (QEs) is given for the three noise conditions (ambient, or quiet; helicopter noise at 60 dBA and at 80 dBA) for the open-ear condition and five hearing protectors (two passive HPDs and three active HPDs).

to 12 trials) following an MRT trial so as to not overlap directly with the primary task.

Dual-task reaction times were better on average for the open ear than for HPD conditions, and word recall was the highest on average for the open-ear condition

CHRISTOPHER J. SMALT, PAUL T. CALAMIA, SHAKTI K. DAVIS, JOSEPH J. LACIRIGNOLA, AND THOMAS F. QUATIERI
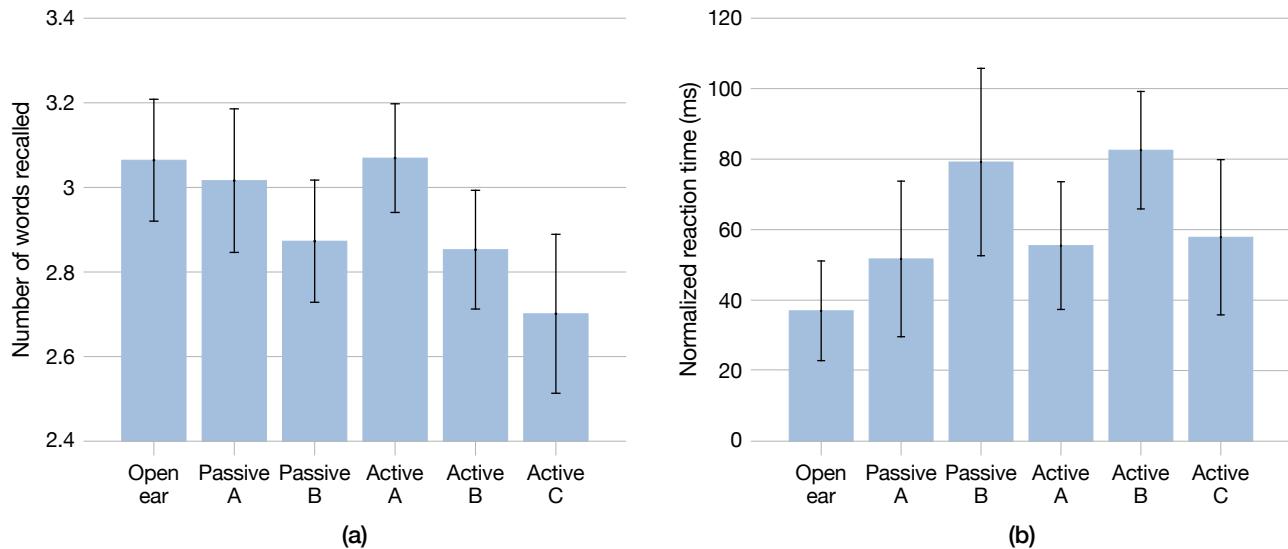
FIGURE 10. The graphs depict the measures of cognitive load induced by the evaluated hearing protection devices (HPDs). Word recall is seen in (a), and visual reaction time is plotted in (b).

(Figure 10a). A two-way ANOVA (with HPD as the fixed factor and subjects as a random factor) was performed on the mean reaction times after subtracting the median baseline time from the same session. This analysis revealed that there was an effect of HPD ($F(4,48) = 3.76$, $p < 0.038$). Post hoc Tukey multiple comparisons ($\alpha = 0.05$) revealed a significant difference between the open condition and the Active B HPD. These results suggest that additional effort required to process speech while wearing hearing protection may have an effect on cognitive resources required to execute the visual reaction-time task.

A separate two-way ANOVA performed on the mean number of words recalled (Figure 10b) revealed that there is no main effect of HPD ($F(4, 45) = 1.65$, $p < 0.178$). However, there is a weak negative correlation ($R = -0.26$, $p < 0.0725$) observed between the reaction time and the number of words recalled across all subjects and hearing-protection devices. This trend indicates that increased visual reaction time and reduced cognitive-load task performance are related, as both are indications of fewer available processing resources.

Computational Models of Auditory Situational Awareness

Computational models of the auditory pathway have been in use since at least the early 1980s and provide a useful tool for testing our understanding of how the auditory system functions and how damage to the cochlea affects both auditory system function and consequently an individual's perception and situational awareness. Functional models based on the physiology or mechanics of the ear, such as the AHAAH model [37], have been adopted by the U.S. Army in MIL-STD-1474E for evaluating impulse-noise exposure and predicting temporary threshold shifts. The models typically share common components: a linear filter for the transfer function of the outer ear and a bandpass filter bank that represents the frequency sensitivity of the cochlea. This filter bank has a number of nonlinear properties that affect the filter width and gain, depending on the acoustic stimulus input level and frequency. While the AHAAH model is designed to predict the risk of hearing damage, Lincoln Laboratory has developed models to predict the effect of hearing damage on auditory performance. A model schematic, which includes a neural component, is shown in Figure 11, where the output is either a measure of speech intelligibility (percentage of words correctly identified) or sound localization (angle error).

Modeling Sound Localization

Because auditory localization performance was effective at discriminating HPDs in the analysis described earlier, a subsequent modeling effort focused on reproducing the human-subject data with a computational

localization model. The purpose of the model is to speed up HPD evaluations, reducing the need for human subject testing and minimizing the time between design cycles. To simulate the original protocol for the model-based analysis, the human listeners were replaced with a GRAS 45CB Acoustic Test Fixture (ATF), a humanoid head-and-shoulders manikin with integrated in-ear microphones (Figure 12). The ATF was fitted with each HPD, and data were collected by playing the AK-47 stimulus twice from each loudspeaker and recording binaural signals with the in-ear microphones. For the evaluation, these binaural signals were used as input to an offline localization algorithm that employed an auditory model to generate angle estimates for each source location. The performance of such a model can be quantified in terms of various error metrics, e.g., the number or percent of front/back confusions, and compared to those same metrics computed on the human-subject data to assess the accuracy of the model-based HPD evaluations.

Once data were collected with the ATF, localization modeling was done in two steps. First, a frontal lateralization angle between –90 degrees and 90 degrees was computed using the approach described by May et al.

[38].[3] Second, the localization process was augmented with a front/back disambiguation step to extend the estimate range to cover the full circle around the listener, i.e., –180 degrees to 180 degrees. Interaural level differences (ILDs, i.e., differences in loudness and frequency distribution between the two ears) are distinct, particularly at high frequencies, for mirrored front-to-back source locations, and are due mainly to direction-dependent shadowing and reflection effects from the pinnae (part of the ear outside the head).

In a preprocessing step, we generated the expected ILDs from an acoustic source located in a series of positions with fine angular resolution around the ATF. Then, during the estimation process, our algorithm followed these steps when provided with binaural signals from an unknown source location:

1. Compute a frontal azimuth estimate θ in the range (–90 degrees to 90 degrees).
2. Compute the ILD spectrum for the input.

---

[3]Localization estimates often are restricted to the frontal hemisphere because some relevant cues, such as interaural time differences, have front/back ambiguities caused by the use of only two ears as sensors.
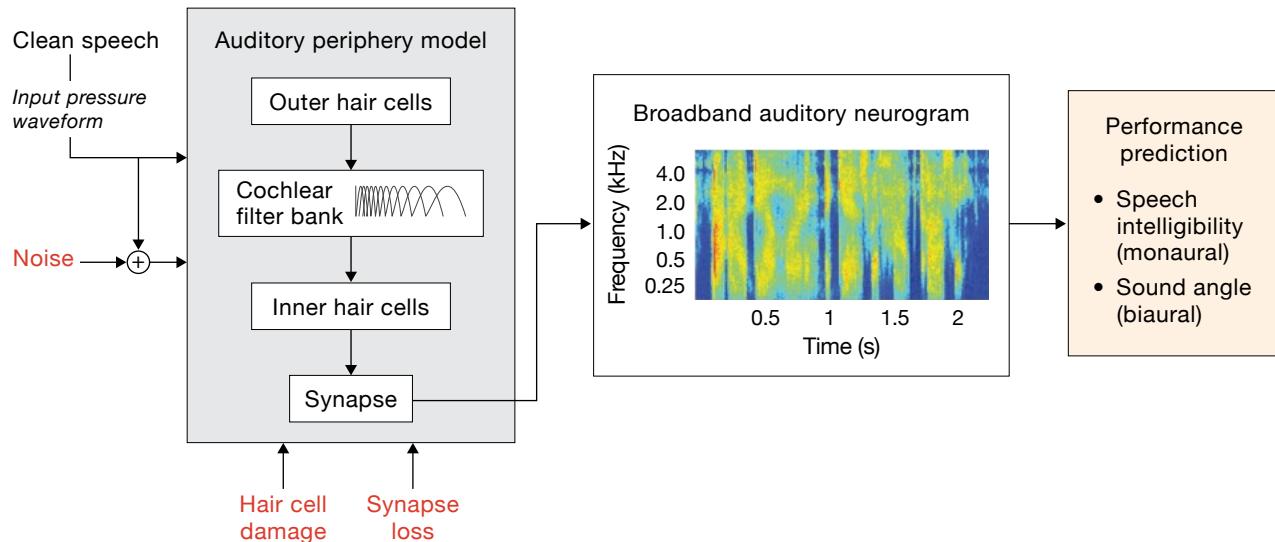
FIGURE 11. The schematic diagram depicts a model to predict speech intelligibility or sound localization performance. The input to the model is a pressure waveform. First the signal is passed to a cochlear periphery model, which is composed of inner and outer hair cells (IHC, OHC) and synapses to generate the auditory nerve output. By simulating the model at various frequencies along the filter bank, a spectrogram-like auditory neurogram output can be generated. Finally, these neurograms can be compared to a reference to estimate intelligibility or across the two ears to estimate the angle of arrival. Other features of the model, shown in red, include the ability to degrade the input with background noise or to simulate hearing loss.

3. Use the azimuth to look up the expected ILD spectra for the corresponding frontal (θ) and rear (180 − θ degrees) positions.

4. Compare the measured ILD spectrum from step 2 to the expected ILD spectra from step 3 and compute the RMS error for each of the two comparisons.

5. Choose the final output azimuth θ (source in front) if the RMS error between the measured ILD spectrum and the expected frontal ILD spectrum is the smaller of the two errors. Otherwise, choose the final output azimuth, 180 degrees to θ (source in back).

Data related to two example angle estimates are shown in Figure 13. In Figure 13a, the measured ILD spectrum closely matches the expected ILD spectrum for a source in the front at 35 degrees azimuth, so the algorithm correctly chooses the frontal location. In Figure 13b, the ILD spectrum is distorted by a hearing protector mounted on the ATF, causing an incorrect choice of the rear position, i.e., a front/back quadrant error similar to one a human listener might make. Preliminary localization modeling results are shown in Figure 14 (the corresponding human subject results are shown in Figure 9). The modeling approach tends to overestimate the number of quadrant errors relative to the number that the human subjects made, but the performance trend among the HPDs is captured.
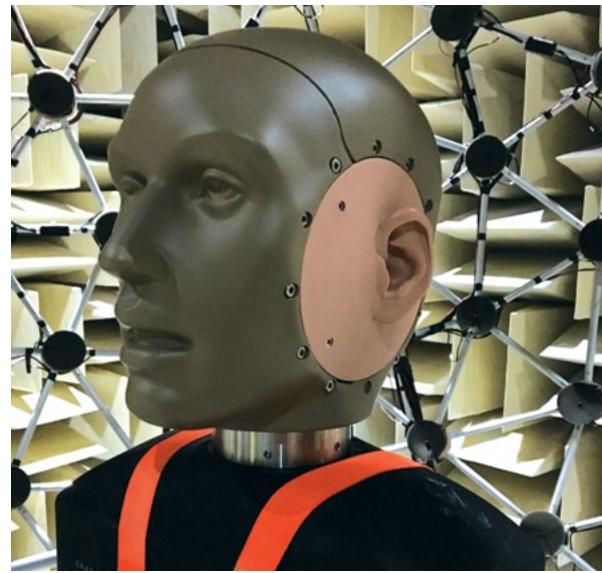
FIGURE 12. The GRAS 45CB Acoustic Test Fixture was used for the model-based hearing-protection device localization assessment.

Modeling Speech Intelligibility

Previously mentioned in this article is the recent discovery that noise exposure can cause a permanent loss of cochlear synapses that is undetectable by a traditional audiogram [6]. The type of synapses that are targeted are those that encode high-level sounds and are associated with low spontaneous rate (LSR) auditory nerve
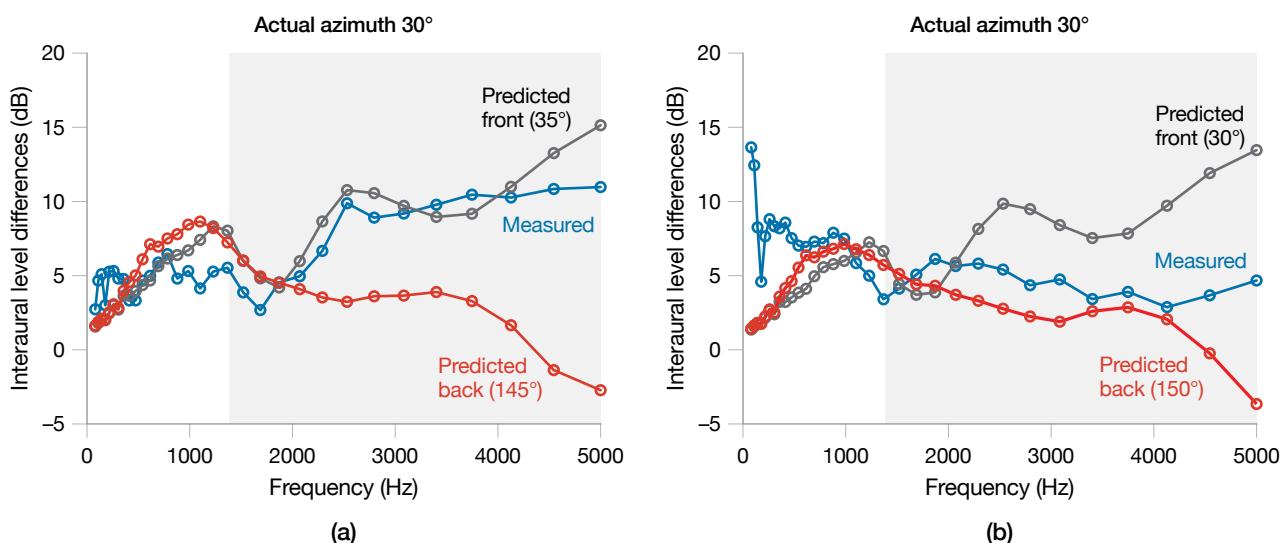
FIGURE 13. The graphs show examples of front/back disambiguation using measured and predicted interaural level differences (ILDs). The data in (a) are for an open-ear test; in (b) are results for an active hearing-protection device (HPD). Note the distortion (blue line in [b]) induced by the HPD.

fibers. This phenomenon is believed to create difficulties for humans to understand speech in noise, although no established clinical or noninvasive technique verifies this assumption. In addition, it is not well understood how LSR loss might interact with the medial olivocochlear reflex (MOCR), a feedback mechanism that controls outer-hair-cell gain. The MOCR is thought to provide protection and signal enhancement for a listener who is in background noise, and it is also used as a predictor of susceptibility to noise injury [39]. To ultimately make predictions of auditory performance, Lincoln Laboratory utilized a computational model of the auditory periphery and cortex to study the effect of LSR auditory-nerve fiber loss and MOCR strength reduction on the cortical representation of speech intelligibility in noise.

We used the auditory-periphery model of Zilany et al. [40, 41] and Smalt et al. [42] to make predictions of auditory-nerve responses to speech stimuli in noise. One hundred auditory-nerve fibers were simulated for each of 32 frequencies along the cochlea between 100 Hz and 8 kHz. The resulting cochlear neurogram, a spectrogram-like output based on auditory-nerve fiber outputs, then became the foundation for measuring the fidelity of speech encoding. At the level of the auditory nerve, this measurement is achieved by computing the Neural Similarity Index (NSIM, [43]), and at the level of the auditory cortex by computing the Spectro-Temporal Modulation Index (STMI, [44]). Both of these techniques essentially compare two copies of the model output: the first is a clean copy of the acoustic target, and the second can be degraded by noise, hearing loss, or both. The strength of the MOCR and the percentage of lost LSR auditory-nerve fibers (i.e., the degree of synaptopathy) can be adjusted in the model, in which the effect on predicted speech intelligibility can be observed. Both STMI and NSIM are normalized measures, where 0 indicates no intelligibility and 1 indicates perfect intelligibility.

Simulations of speech intelligibility using the neurogram revealed that both the MOCR and LSR loss have minimal effect on STMI speech intelligibility predictions in quiet (ΔSTMI < 0.04). In background noise, however, the effect of the MOCR and LSR loss was more significant (ΔSTMI$_{MOCR}$ = 0.23, ΔSTMI$_{LSR}$ = 0.07). This effect can also be observed qualitatively in the neurogram images (Figure 15). In these simulations, a complete loss of the LSR fibers in the population response was used to
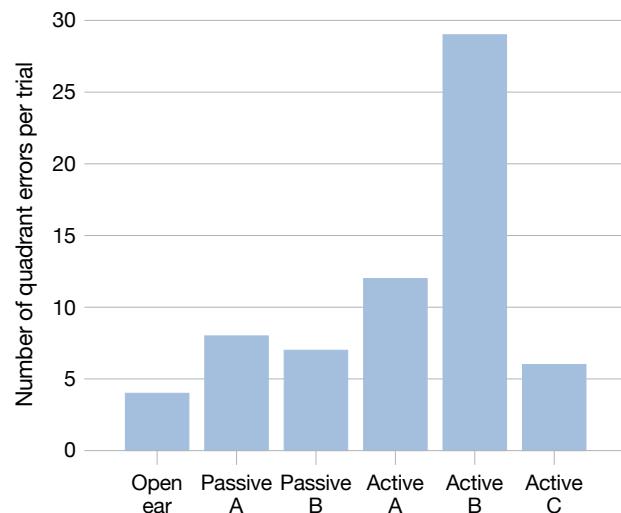
FIGURE 14. Preliminary results for tests using a localization model are similar to those derived from human-subject testing (Figure 9), as indicated by the average number of quadrant errors (QEs).

indicate the maximum possible damage that could occur. Preliminary results on a 50-subject human test showed a better correlation with speech intelligibility in noise using STMI and MOCR (r = 0.38) as compared to the audiogram alone (R = 0.3).

These results suggest that while LSR loss can have some impact directly on neural signal fidelity in noisy environments, efferent feedback such as the MOCR has a stronger effect on speech perception. One possible explanation for this observation is that the LSR fibers are indirectly responsible for reducing MOCR strength, which in turn causes communication difficulties. This hypothesis is supported by the fact that LSR fibers typically encode louder sounds, and it is known that they drive the input to the middle-ear reflex. Further simulations using the auditory-nerve model may help researchers understand human susceptibility to noise-induced hearing loss, the impact of damage to low spontaneous rate auditory-nerve fibers, and the health and performance risks of noise exposure to the warfighter.

Discussion
Implications of Compromised Hearing
One outstanding question related to auditory health and situational awareness is how degradations caused by hearing loss or hearing protection affect the warfighter's

STMI = Spectro-Temporal Modulation Index
NSIM = Neural Similarity Index
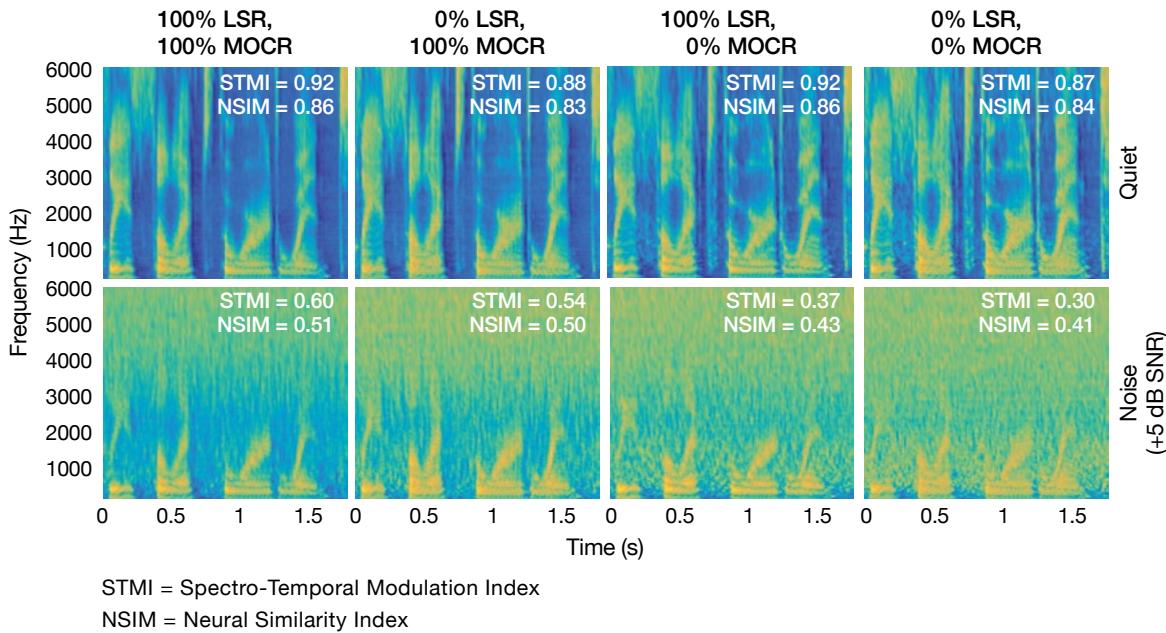
FIGURE 15. The neurograms of speech in quiet and in broadband background noise show results for a healthy cochlea, 100 percent low spontaneous rate (LSR) auditory nerve fibers, and 100 percent medial olivocochlear reflex (MOCR) strength. The signal-to-noise ratio of the neurogram is reduced significantly because of synaptic loss (LSR and efferent feedback loss [MOCR]) for speech in background noise but not in quiet. This simulation may explain why the predominant hearing issue people have is listening in background noise; it is likely to be a result of synaptic loss in the cochlear nerve from noise exposure.

operational performance. Put another way, at what point does a change in detection threshold, or localization accuracy, or speech intelligibility, or cognitive load, or some combination of these, become a hindrance to mission success or survivability? This complex issue has been studied only to a limited degree thus far in laboratory and field experiments, and through surveys of active-duty soldiers.

Field Studies

Casali et al. [45] studied the effects of three HPDs on in-field performance of reconnaissance and raid missions through subjective and objective measures of auditory detection and identification of threats and through communication with participants in the experiments. Casali et al. found varied performance over missions and ratings and no indication of an optimal HPD among the ones included in their experiments. While they did receive subjective feedback from participants about aspects of situational awareness, such as localization abilities, they did not provide measurements of HPD performance.

Clasing and Casali [2] studied detection and identification of auditory threats (gunshot, spoken Arabic, and weapons preparation) with five HPDs and open ear; they found performance varied among the HPDs, with occasional performance improvements (from one HPD) but mainly detriments. Again, no measurements of HPD performance were provided.

Talcott et al. [46] studied azimuthal localization of and response time to gunshots with four HPDs and the open ear. They found that localization was generally worse with all HPDs (none preserved "normal" performance), but they did not comment on the impact any of the performance degradations might have.

Sheffield et al. [3] used hearing-loss simulation systems to study the effect of hearing loss on operational performance in a dismounted combat scenario (a last-man-standing paintball competition). While their results indicated a reduction in offensive effectiveness with increased hearing impairment but little effect on survivability, they did not map their simulated hearing profiles to metrics such as localization or intelligibility.

Laboratory Studies

Peters and Garinther [47] studied the effects of degraded speech intelligibility on the performance of tank operators in a simulator. Speech intelligibility was modulated by electronically distorting the speech in the simulator, and the performance of two-man crews was quantified in terms of mission time, completion, and error, as well as gunner accuracy. All performance metrics other than accuracy were shown to degrade with reduced intelligibility.

Similarly, Mentel et al. [48] evaluated the relationship between speech intelligibility and operational performance in a simulation of an Aegis Combat System Command Information Center. Using a modified rhyme test, they found performance, in terms of the percentage of successfully accomplished tasks, decreased significantly when modified rhyme test scores fell below 65 percent correct. Neither of these studies involved HPDs as a variable, although our results suggest that noise level is more important than HPD choice when assessing speech intelligibility.

Surveys

By surveying military personnel with infantry or combat-support roles with a questionnaire, Semeraro et al. [49] identified nine mission-critical auditory tasks (seven communication, one detection, and one localization) to be included in future tests of auditory fitness. However, while they asked the participants to comment on the consequences of poor performance for each task, they neither inquired about nor tried to define acceptable performance levels.

Auditory Fitness for Duty

All of these studies can be classified as attempts to evaluate what is known as auditory fitness for duty, or "the possession of hearing abilities sufficient for safe and effective job performance" [50]. Whether auditory performance is limited by hearing impairment or the use of HPDs, the effects of such limitations should be linked to job performance more directly. Such study of this linkage will help determine the suitability of people with or without hearing protection or enhancement for the tasks to which they are assigned. Efforts to help define tests for auditory fitness for duty have been discussed for military [49, 51] and law-enforcement [52] personnel, but significant work remains to be done in this area to

understand the relative contributions of detection, localization, speech intelligibility, cognitive load, and other factors and to define appropriate testing protocols.

Continuing Research

Auditory health is a significant concern within the DoD and VA because typical high-noise military environments, whether associated with training or operational activities, are responsible for hearing loss and tinnitus affecting large numbers of active-duty service personnel and veterans. Reductions in hearing ability are known to affect the operational performance of a soldier, both on and off the battlefield, in ways that must be better understood to determine the fitness of a person for a specific job and to develop devices that provide ample hearing protection. Lincoln Laboratory researchers are involved in research to improve dosimetry devices that address the complexity of military noise environments comprising continuous and impulse noise, often with moving sound sources and listeners. We have investigated damage-risk metrics needed to estimate the likelihood and severity of noise-induced hearing injuries, and we have developed computational models of the auditory pathway to evaluate detrimental HPD-induced effects on auditory perceptual tasks, such as localizing sounds and understanding speech. In the future, we will continue to seek advancements for evaluating and maintaining auditory health that will lead to improved warfighter performance and reduced hearing-related disabilities. ■

References

1. Veterans Benefits Administration, "Annual Benefits Report: Fiscal Year 2016." Washington, D.C.: U.S. Department of Veterans Affairs, 2016.

2. J.E. Clasing and J. Casali, "Auditory Situation Awareness: Effects of Enhanced Hearing Protection Devices and TCAPS on In-Field Detection and Identification of Three Military Threat Signatures," paper in *Proceedings of the 41st International Congress and Exposition on Noise Control Engineering (INTER-NOISE 2012)*, vol. 5, 2012, pp. 3818–3829.

3. B. Sheffield, D. Brungart, J. Tufts, and J. Ness, "The Relationship Between Hearing Acuity and Operational Performance In Dismounted Combat," paper in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 59, no. 1, 2015, pp. 1346–1350.

4. S.M. Abel, "Barriers to Hearing Conservation Programs in Combat Arms Occupations," *Aviation, Space, and Environmental Medicine*, vol. 79, no. 6, 2008, pp. 591–598.

5. V. Tepe, C. Smalt, J. Nelson, T. Quatieri, and K. Pitts, "Hidden Hearing Injury: The Emerging Science and Military Relevance of Cochlear Synaptopathy," *Military Medicine*, vol. 182, no. 9, 2017, pp. e1785–e1795.

6. S.G. Kujawa and M.C. Liberman, "Adding Insult to Injury: Cochlear Nerve Degeneration after 'Temporary' Noise-Induced Hearing Loss," *Journal of Neuroscience*, vol. 29, no. 45, 2009, pp. 14077–14085.

7. H.W. Lin, A.C. Furman, S.G. Kujawa, and M.C. Liberman, "Primary Neural Degeneration in the Guinea-Pig Cochlea after Noise-Induced Threshold Shift," *Journal of the Association for Research in Otolaryngology*, vol. 12, no. 5, 2011, pp. 605–616.

8. A.C. Furman, S.G. Kujawa, and M.C. Liberman, "Noise-Induced Cochlear Neuropathy Is Selective for Fibers with Low Spontaneous Rates," *Journal of Neurophysiology*, vol. 110, no. 3, 2013, pp. 577–586.

9. J.B. Jensen, A.C. Lysaght, M.C. Liberman, K. Qvortrup, and K.M. Stankovic, "Immediate and Delayed Cochlear Neuropathy after Noise Exposure in Pubescent Mice," *PLoS ONE*, vol. 10, no. 5: e0125160, 2015.

10. Undersecretary of Defense for Acquisition, Technology, and Logistics, "Hearing Conservation Program," DoD Instruction 6055.12. Washington, D.C.: U.S. Department of Defense, 2010.

11. American Institute of Biological Sciences, "Peer Review of Injury Prevention and Reduction: Impulse Noise Injury Models." Frederick, Md.: U.S. Army Medical Research and Materiel Command and Military Operational Medicine Research Program, 2010, pp. 1–15.

12. R.P. Hamernik, D. Henderson, J.J. Crossley, and R.J. Salvi, "Interaction of Continuous and Impulse Noise: Audiometric and Histological Effects," *The Journal of the Acoustical Society of America*, vol. 55, no. 1, 1974, pp. 117–121.

13. R.P. Hamernik, W.A. Ahroon, and K.D. Hsueh, "The Energy Spectrum of an Impulse: Its Relation to Hearing Loss," *The Journal of the Acoustical Society of America*, vol. 90, no. 1, 1991, pp. 197–204.

14. G.R. Price, "Impulse Noise Hazard as a Function of Level and Spectral Distribution," in R.J. Salvi, D. Henderson, R.P. Hamernik, and V. Colletti, eds., *Basic and Applied Aspects of Noise-Induced Hearing Loss*. Boston: Springer, 1986.

15. U.S. Department of Defense, "Department of Defense Design Criteria Standard: Noise Limits," MIL-STD-1474E. Washington, D.C.: U.S. Department of Defense, 2015.

16. G.R. Price, "Validation of the Auditory Hazard Assessment Algorithm for the Human with Impulse Noise Data." T*he Journal of the Acoustical Society of America*, vol. 122, no. 5, 2007, pp. 2786–2802.

17. J.T. Kalband and G.R. Price, "Mathematical Model of the Ear's Response to Weapons Impulses," Conference Paper ARL-RP-0521. Aberdeen Proving Ground, Md.: Army Research Lab, Human Research and Engineering Directorate, 2015.

18. B. Zagadou, P. Chan, K. Ho, and D. Shelley, "Impulse Noise Injury Prediction Based on the Cochlear Energy," *Hearing Research*, vol. 342, 2016, pp. 23–38.

19. H.G. Jones, N.T. Greene, and W.A. Ahroon, "Assessment of Middle Ear Function during the Acoustic Reflex Using Laser-Doppler Vibrometry," Technical Report USAARL-2017-16. Fort Rucker, Ala.: U.S. Army Aeromedical Research Laboratory, 2017.

20. H.G. Jones, N.T. Greene, and W.A. Ahroon, "Assessment of Middle Ear Function during the Acoustic Reflex Using Laser-Doppler Vibrometry." Technical Report USAARL-2017-16. Fort Rucker, Ala.: Army Aeromedical Research Laboratory, 2017.

21. W. Qiu, R.P. Hamernik, and B. Davis, "The Kurtosis Metric as an Adjunct to Energy in the Prediction of Trauma from Continuous, non-Gaussian Noise Exposures," *The Journal of the Acoustical Society of America*, vol. 120, no. 6, 2006, pp. 3901–3906.

22. G.S. Goley, W.J. Song, and J.H. Kim, "Kurtosis Corrected Sound Pressure Level as a Noise Metric for Risk Assessment of Occupational Noises," *The Journal of the Acoustical Society of America*, vol. 129, no. 3, 2011, pp. 1475–1481.

23. W. Qiu, R.P. Hamernik, and R.I. Davis, "The Value of a Kurtosis Metric in Estimating the Hazard to Hearing of Complex Industrial Noise Exposures," *The Journal of the Acoustical Society of America*, vol. 133, no. 5, 2013, pp. 2856–2866.

24. P. Sun, J. Qin, and W. Qiu, "Development and Validation of a New Adaptive Weighting for Auditory Risk Assessment of Complex Noise," *Applied Acoustics*, vol. 103, 2016, pp. 30–36.

25. E.A. Shaw and M.M. Vaillancourt, "Transformation of Sound-Pressure Level from the Free Field to the Eardrum Presented in Numerical Form," *The Journal of the Acoustical Society of America*, vol. 78, no. 3, 1985, pp. 1120–1123.

26. "Specification for Personal Noise Dosimeters," ANSI/ASA S1.25-1991. New York: American National Standards Institute, 1991.

27. C.A. Kardous and R.D. Willson, "Limitations of Using Dosimeters in Impulse Noise Environments," *Journal of Occupational and Environmental Hygiene*, vol. 1, no. 7, 2004, pp. 456–462.

28. C.A. Kardous, R.D. Willson, and W.J. Murphy, "Noise Dosimeter for Monitoring Exposure to Impulse Noise," *Applied Acoustics*, vol. 66, no. 8, 2005, pp. 974–985.

29. C.A. Kardous and P.B. Shaw, "Evaluation of Smartphone Sound Measurement Applications," *The Journal of the Acoustical Society of America*, vol. 135, no. 4, 2014, pp. EL186–EL192.

30. "American National Standard Methods for the Measurement of Insertion Loss of Hearing Protection Devices in Continuous or Impulsive Noise Using Microphone-in-Real-Ear or Acoustic Test Fixture Procedures," ANSI/ASA S12.42. New York: American National Standards Institute, 2010.

31. D.L. Johnson, "Blast Overpressure Studies with Animals and Man: Walk-up Study." Fort Detrick, Frederick, Md.: U.S. Army Medical Research and Development Command, 1993.

32. R. Reddy, D. Welch, S. Ameratunga, and P. Thorne, "Development of the Hearing-Protection Assessment (HPA-2) Questionnaire," *Occupational Medicine*, vol. 64, no. 3, 2014, pp. 198–205.

33. H.L. Gallagher, R.L. McKinley, M.A. Theis, B.J. Swayne, and E.R. Thompson, "Performance Assessment of Passive Hearing Protection Devices," Technical Report AFRL-RH-WP-TR-2014-0148. Wright-Patterson Air Force Base: Air Force Research Laboratory, 2014.

34. A.D. Brown, B.T. Beemer, N.T. Greene, T. Argo IV, D. Meegan, and D.J. Tollin, "Effects of Active and Passive Hearing Protection Devices on Sound Source Localization, Speech Recognition, and Tone Detection," *PloS One*, vol.10, no. 8: e0136568, 2015.

35. A.S. House, C. Williams, M.H.L. Hecker, and K.D. Kryter, "Psychoacoustic Speech Tests: A Modified Rhyme Test," *The Journal of the Acoustical Society of America*, vol. 35, no. 11, 1963, pp. 1899–1899.

36. B.W. Hornsby, "The Effects of Hearing-Aid Use on Listening Effort and Mental Fatigue Associated with Sustained Speech Processing Demands," *Ear and Hearing*, vol. 34, no. 5, 2013, pp. 523–534.

37. G.R. Price and J.T. Kalb, "Insights into Hazard from Intense Impulses from a Mathematical Model of the Ear," *The Journal of the Acoustical Society of America*, vol. 90, no. 1, 1991, pp. 219–227.

38. T. May, S. van de Par, and A. Kohlrausch, "A Probabilistic Model for Robust Localization Based on a Binaural Auditory Front-End," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 1, 2011, pp.1–13.

39. S.F. Maison and M.C. Liberman, "Predicting Vulnerability to Acoustic Injury with a Noninvasive Assay of Olivocochlear Reflex Strength," *Journal of Neuroscience*, vol. 20, no. 12, 2000, pp. 4701–4707.

40. M.S. Zilany, I.C. Bruce, P.C. Nelson, and L.H. Carney, "A Phenomenological Model of the Synapse Between the Inner Hair Cell and Auditory Nerve: Long-Term Adaptation with Power-Law Dynamics," *The Journal of the Acoustical Society of America*, vol. 126, no. 5, 2009, pp. 2390–2412.

41. M.S.A. Zilany, I.C. Bruce, and L.H. Carney, "Updated Parameters and Expanded Simulation Options for a Model of the Auditory Periphery," *The Journal of the Acoustical Society of America*, vol. 135, no. 1, 2014, pp. 283–286.

42. C.J. Smalt, M.G. Heinz, and E.A. Strickland, "Modeling the Time-Varying and Level-Dependent Effects of the Medial Olivocochlear Reflex in Auditory Nerve Responses," *Journal of the Association for Research in Otolaryngology*, vol. 15, no. 2, 2014, pp. 159–173.

43. A. Hines and N. Harte, "Speech Intelligibility Prediction Using a Neurogram Similarity Index Measure," *Speech Communication*, vol. 54, no. 2, 2012, pp. 306–320.

44. M. Elhilali, T. Chi, and S.A. Shamma, "A Spectro-Temporal Modulation Index (STMI) for Assessment of Speech Intelligibility," *Speech Communication*, vol. 41, nos. 2–3, 2003, pp. 331–348.

45. J.G. Casali, W.A. Ahroon, and J.A. Lancaster, "A Field Investigation of Hearing Protection and Hearing Enhancement in One Device: For Soldiers Whose Ears and Lives Depend upon It," *Noise and Health*, vol. 11, no. 42, 2009, pp. 69–90.

46. K.A. Talcott, J.G. Casali, J.P. Keady, and M.C. Killion, "Azimuthal Auditory Localization of Gunshots in a Realistic Field Environment: Effects of Open-Ear versus Hearing Protection-Enhancement Devices (HPEDs), Military Vehicle Noise, and Hearing Impairment," *International Journal of Audiology*, vol. 51, suppl. 1, 2012, pp. S20–S30.

47. L.J. Peters and G.R. Garinther, "The Effects of Speech Intelligibility on Crew Performance in an M1A1 Tank Simulator," Technical Memorandum 11-90. Aberdeen Proving Ground, Md.: U.S. Army Human Engineering Laboratory, 1990.

48. K. Mentel, J. Ziriax, J. Dachos, A. Salunga, H. Turner, B. Sheffield, and D. Brungart, "The Relationship Between Speech Intelligibility and Operational Performance in a Simulated Naval Command Information Center," paper in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 57, no. 1, 2013, pp. 1204–1208.

49. H.D. Semeraro, Z.L. Bevis, D. Rowan, R.M. van Besouw, and A.J. Allsopp, "Fit for the Frontline? Identification of Mission-Critical Auditory Tasks (MCATs) Carried out by Infantry and Combat-Support Personnel," *Noise and Health*, vol. 17, no. 75, 2015, p. 98–107.

50. J.B. Tufts, K.A. Vasil, and S. Briggs, "Auditory Fitness for Duty: A Review," *Journal of the American Academy of Audiology*, vol. 20, no. 9, 2009, pp. 539–557.

51. D.S. Brungart, "20Q: Auditory Fitness for Duty," *Audiology*, online article 12528, 2014, retrieved from http://www.audiologyonline.com on 21 April 2016.

52. V. Vaillancourt, C. Laroche, C. Giguère, M.A. Beaulieu, and J.P. Legault, "Evaluation of Auditory Functions for Royal Canadian Mounted Police Officers," *Journal of the American Academy of Audiology*, vol. 22, no. 6, 2011, pp. 313–331.

About the Authors

Christopher J. Smalt is a technical staff member in the Human Health and Performance Systems Group at Lincoln Laboratory. His current work focuses on computational auditory modeling, specifically on mechanisms of hearing damage and the effect of noise exposure on hearing and cognitive performance. He and a team at the Laboratory have developed on-body and in-ear noise monitoring technologies suitable for industrial and tactical environments. His other research interests include 3D virtual audio, cognitive load, and auditory electrophysiology. Prior to joining the Laboratory in 2013, he was a doctoral student in the Department of Electrical and Computer Engineering at

Purdue University. There he modeled efferent feedback mechanisms in the auditory pathway and studied adaptation to an acoustic simulation of cochlear implants through fMRI imaging, supported in part by a Predoctoral Fellowship Training Grant from the National Institutes for Health/National Institute on Deafness and Other Disorders. He holds a bachelor's degree in computer engineering from Clarkson University and earned both master's and doctoral degrees in electrical engineering from Purdue University. He also studied trumpet performance at the Crane School of Music at State University of New York Potsdam, and is a member of the Longwood Symphony Orchestra in Boston, which raises money for local medical charities.

Paul T. Calamia was a member of the technical staff in the Bioengineering Systems and Technologies Group. His research interests include auditory health, performance, and perception, as well as audio/acoustic signal processing and modeling. He previously worked on undersea sensing problems with the Advanced Sensor Techniques Group and the Advanced Undersea Systems and Technology Group. Prior to joining Lincoln Laboratory in 2009, he was an assistant professor in the Graduate Program in Architectural Acoustics at Rensselaer Polytechnic Institute, where his research focused on computational models for room acoustics. He holds a bachelor's degree in mathematics from Duke University, a master's degree in electrical and computer engineering from the Engineering Acoustics Program at the University of Texas at Austin, and a doctorate in computer science from Princeton University.

Shakti K. Davis is a member of the technical staff in the Human Health and Performance Systems Group. Her research interests include sensor and signal processing algorithm development with the aims of protecting the warfighter against injury and enhancing performance capabilities. Since 2015, she has supported a variety of signal processing research efforts, including characterizing exposure to hazardous military noise through the use of body-worn acoustic sensors; developing algorithms to estimate risk factors of musculoskeletal injury by using foot-mounted biomechanical sensors; and classifying methods for early warning of virus exposure with high-resolution noninvasive physiological sensors. She joined Lincoln Laboratory in 2006 as part of the Sensor Systems Division, where she worked on detection and classification algorithms for airborne radar systems. She holds a bachelor's degree from New Mexico State University and a doctoral degree from the University of Wisconsin, Madison in electrical engineering.

Joseph J. Lacirignola is a technical staff member in the Counter–Weapons of Mass Destruction Systems Group. His research interests include sensor, device, and novel detection development capabilities. He joined the Laboratory in 2003 and worked on CANARY (i.e., Cellular Analysis and Notification of Antigen Risks and Yields) and related biosensing hardware platforms. Throughout his career, he has been involved in numerous hardware programs spanning biological and chemical detection, wearable sensing, novel 3D-printing hardware, multimodal detection systems, and the development of unmanned aerial and ground vehicles. He has authored several journal articles, conference presentations, and book chapters and holds more than 10 patents. He holds a bachelor's degree in biology from Syracuse University, bachelor's degrees in mechanical engineering and mechanical design from Wentworth Institute of Technology, and a master's degree in molecular and cell biology from Brandeis University.

Thomas F. Quatieri, a senior member of the technical staff in the Human Health and Performance Systems Group, is involved in bridging human language and bioengineering research and technologies. Within this group, he has initiated and developed major R&D and technology transition programs in speech and auditory signal processing and neuro-biophysical modeling with application to detection and monitoring of neurological, neurotraumatic, and cognitive stress conditions. He has been an author on more than 200 publications, holds 11 patents, and authored the textbook *Discrete-Time Speech Signal Processing: Principles and Practice*. He also holds a faculty appointment in the Harvard-MIT Speech and Hearing Bioscience and Technology Program. He developed the MIT graduate course Digital Speech Processing and is active in advising graduate students on the MIT and Harvard campuses. He is a recipient of four IEEE Transactions Best Paper Awards and the 2010 MIT Lincoln Laboratory Best Paper Award. He led the Lincoln Laboratory team that won the 2013 and 2014 AVEC Depression Challenges and the 2015 MIT Lincoln Laboratory Team Award for their work on vocal and facial biomarkers. He has served on the IEEE Digital Signal Processing Technical Committee, the IEEE Speech and Language Technical Committee, and the IEEE James L. Flanagan Speech and Audio Awards Committee. He has served on National Institutes of Health and National Science Foundation panels, been an associate editor for the *IEEE Transactions on Signal Processing*, and is an associate editor of *Computer, Speech, and Language*. He has been an invited speaker at conferences, workshops, and meetings, most recently as a keynote speaker at the 2019 Speech Science and Technology Conference. He is a Fellow of the IEEE and a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, ICSA, SFN, ARO, and ASA. He holds a bachelor's degree (summa cum laude) from Tufts University, and master's and doctoral degrees from MIT.

Model and Personal Sensor for Metabolic Tracking and Optimization

Lawrence M. Candell, Gary A. Shaw, and Kyle J. Thompson

Lincoln Laboratory has developed a novel metabolic fuel model and low-cost breath sensor for measuring, tracking, and enhancing metabolism. The model can predict key metabolic state parameters, including blood glucose levels, available glycogen stores, nutrient substrate utilization, and fat accumulation or depletion, for a given diet and activity profile. It can also predict healthy metabolic responses to a variety of dietary interventions and the onset of medical conditions, such as type 2 diabetes and excessive fat accumulation. The model suggests that the measurement of key metabolic parameters, such as respiratory quotient and energy expenditure, can provide insight into metabolic health and improvements in athletic performance and endurance.

>>

At the cellular level, metabolism refers to living cells' physical and chemical reactions that produce the energy required for life. These reactions are generally categorized as catabolic (breaking down large molecules to release energy) or anabolic (synthesizing complex molecules to store energy). Humans' very existence is intimately tied to the successful operation of these cellular metabolic processes. Despite the importance of metabolism, most people possess little insight into its function and often have a vague understanding of how dietary intake and activities impact metabolism. Historically speaking, metabolism and nutrition are relatively immature fields of study, and much of what is understood about the complex regulation of cellular metabolism and its relationship to nutrition has been gleaned in just the last half century.

The metabolic processes responsible for converting dietary macronutrients into the chemical energy needed to power biochemical processes were not fully understood until early in the twentieth century, when physiologists Hans Krebs and Fritz Albert Lipmann discovered the citric acid cycle [1]. As understanding of the science behind nutrition and metabolism has evolved, so has the ability to identify and treat metabolic disorders related to dietary imbalances. Despite progress in understanding metabolism, controversies still exist regarding its basic principles, such as the virtues of a low-carbohydrate versus a low-fat diet. Furthermore, even the settled science is slow to influence public policy and health care partly because nutrition education has historically played a minor role in medical training [2]. In a 2006 study,
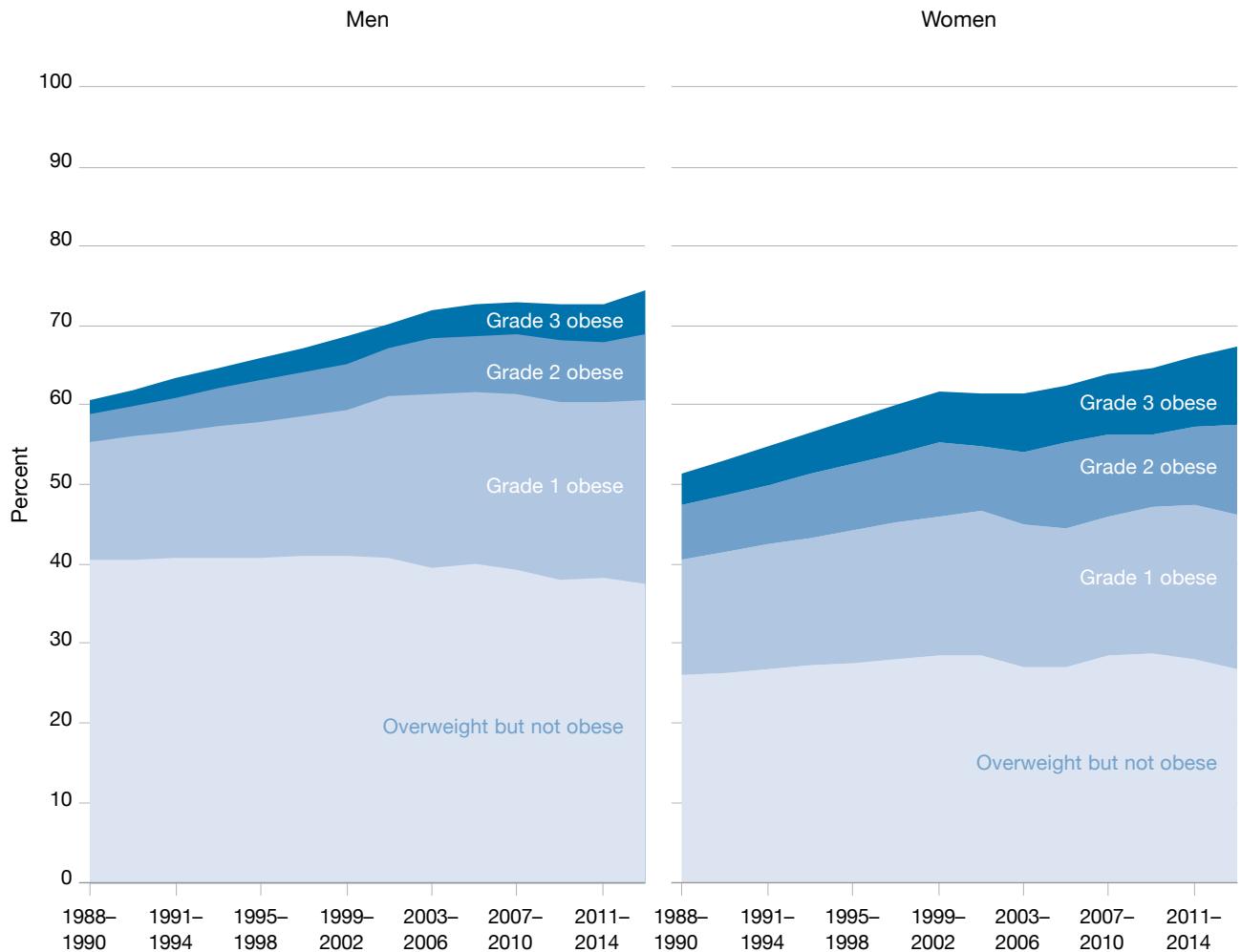
FIGURE 1. The rate of obesity has been increasing in the U.S. population for decades. There has been no consensus regarding how to reverse the trend. Grade 3 obesity is defined as having a body mass index (BMI) > 40. Grade 2 obesity is 40 > BMI > 35; Grade 1 obesity is 35 > BMI > 30; overweight but not obese is 30 > BMI > 25; and healthy is 24.9 > BMI > 18.5 [3].

less than one-third of U.S. medical schools were found to provide adequate nutrition education in their medical training curricula [4].

The relationships between diet, metabolism, socio-economic status, and debilitating medical conditions, such as type 2 diabetes, are still subjects of active research and debate [5–10]. The lack of consensus is evident from the multitude of different, often antithetical (e.g., low-fat versus low-carbohydrate) diet plans, and degraded or dysfunctional metabolisms among a significant fraction of the U.S. population. According to the Centers for Disease Control and Prevention, in 2016, nearly half of the adults in the United States were either prediabetic (86 million) or diabetic (29 million) [11]. In

addition, more than 35 percent of American adults and close to 17 percent of children were obese. As indicated in Figure 1, the percentage of the population affected by metabolic diseases continues to grow [12]. While there are many postulated factors driving the rapid (within one to two generations) growth in obesity, there is neither consensus regarding the root cause nor success in stemming this epidemic.

Metabolic diseases also represent a national security issue across all service branches and phases of a military career, including recruitment, training, deployment, and post-separation from the service. According to a 2015 report, one in three young adults of recruitment age is unqualified for military service because of excessive
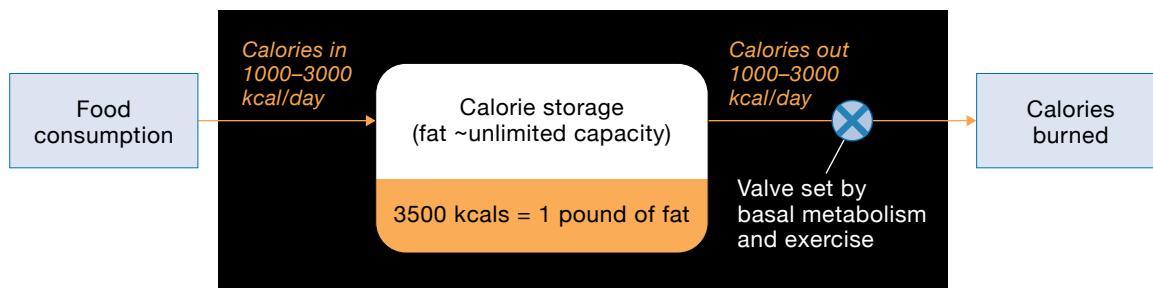
FIGURE 2. The calories-in/calories-out metabolic model helps explain the method of weight gain/loss and is the basis for many weight management strategies. Calories are measured in kilocalories (kcals).

body weight or weight-related disorders [13, 14]. During training and deployment, metabolic fitness levels impact mission readiness, and optimizing a soldier's performance requires matching nutrition to individual metabolic needs [15]. During the later stages of a military career and retirement, avoiding obesity and type 2 diabetes is a concern for the service organizations and aging soldiers. Currently, 15 percent of the patients in Veterans Health Administration hospital care are diabetic, and the Department of Defense spends more than $4 billion per year treating weight-related diseases [16].

The U.S. military has an interest in comprehensive metabolic measurement and tracking systems for optimizing the performance of soldiers under demanding physical conditions and for maintaining soldiers' metabolic health and wellness. It is in this context that Lincoln Laboratory has been developing a novel metabolic model to predict how the body will respond to changes in dietary intake and activity level. The body's response can vary from person to person, with training, and over time; this variability requires measuring individual metabolic parameters to fully personalize the model.

What Is a Metabolic Model, and Why Do We Need One?

To improve public health and optimize physical/mental performance, we must first understand how metabolism varies from individual to individual and how it responds to diet and exercise. An engineering approach to problem solving and optimization often begins with the creation of a system-level model, in which system level implies a model that possesses sufficient fidelity to represent the behavior(s) of interest while keeping complexity to a minimum. For example, if the problem of interest

involved modeling the flight dynamics of an aircraft, a parsimonious system model would not include the inner workings of the engines or the individual integrated circuits. Instead, it would include only the salient performance parameters, such as available thrust, lift, or drag, and the control authority to influence all of these factors while omitting the mechanism by which the control was achieved or implemented. To help place our unique metabolic model in context, we review two extremes of metabolic modeling complexity that are frequently invoked in the discussion of metabolism and metabolic health maintenance.

Calories-In/Calories-Out Model

From the general population and from many health professionals, the most common answer to the question "What is the most important metric for managing body weight?" is calories, with the understanding that energy balance is needed for weight maintenance, and an energy deficit is needed for weight loss [17–19]. The calories-in/calories-out (CICO) model, illustrated in Figure 2, captures the essence of the first law of thermodynamics, which states that energy transforms from one form to another but is never created or destroyed.

The model accounts for the fact that if more calories are ingested than expended, the excess calories not excreted are stored (e.g., as glycogen in the liver or muscle tissue or as fat in adipose tissue), resulting in an increase in body weight. Using the CICO model as a base, many weight management strategies emphasize tracking the caloric content of foods eaten and the calories expended while exercising. These weight management plans sometimes include measuring an individual's resting metabolic rate to quantify, rather than estimate, basal

metabolic energy needs and to create a more accurate calorie-restricted diet and exercise regimen to achieve the desired energy deficit for weight loss [20].

While the physics of the CICO model are technically correct, the model is insufficient to provide insight into the complex human metabolism. In particular, the CICO model makes no distinction regarding dietary macro-nutrient composition (fat versus carbohydrate versus protein) or the homeostatic mechanisms inciting hunger or satiety and is thus unable to explain phenomena such as metabolic set points (a preferred or predetermined body weight regulated by a feedback control mechanism) or satiety [21, 22]. For example, a one-liter soda and a small salad both provide about 450 calories. However, the body's response to these two meals is quite different in terms of how the calories are processed, how quickly they become available to meet energy needs, how they impact blood glucose levels, and how much satiety they produce. The CICO model is far too simple to provide useful insight into how the different types of calories[1] we ingest impact metabolic health and performance.

Cellular-Level Models and Utility

At the other extreme of complexity are cellular-level models that seek to represent the microscale processes contributing to metabolism before building them into a holistic human model [23, 24]. However, given that there are tens of trillions of interdependent cells in the human body, micro and cellular models are too detailed for efficient scaling to a whole-body model. What is needed is a model with complexity falling between the CICO model, which fails to account for different macro-nutrients, and cellular-level models that don't scale well to system-level simulation.

An Engineering Approach to Goldilocks Metabolic Modeling

The genesis of the system model of metabolism intro-duced here is strongly influenced by the observation that the human body prioritizes maintaining blood glucose concentration within a narrow range [25]. For an average-size adult, controlling blood sugar levels to

[1]Note that a food calorie, denoted with an uppercase C, is equivalent to 1,000 of the calories defined in physics and chemistry. To transition between discussions of food calories and energy expenditure, we will employ the physics definition of a calorie, with the understanding that one kilocalorie (kcal) is equivalent to one food calorie.
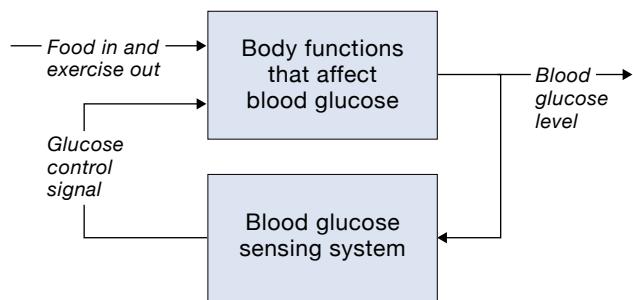
FIGURE 3. This generic blood glucose control model exemplifies a feedback control system and considers the input of food into the body, the energy demands of exercise, and the current blood glucose level to generate a control signal that regulates metabolic homeostasis.

a nominal value of 100 milligrams/deciliter (mg/dL) translates to a control objective of maintaining about five grams of glucose within the body's entire circu-lating blood volume. The fact that this level stays fairly constant, regardless of whether carbohydrates are plentiful (300+ grams of carbohydrate per day) or nonexistent (zero grams per day), does not happen by chance. Blood glucose levels must be actively controlled through feedback, and the body employs a system of mechanisms to monitor blood glucose level, signal that the level is out of range, and actively control or correct glucose levels in response to these signals. It is this system control that is captured in our model and drives the observed metabolic behavior.

Key Elements of a Metabolic Fuel Model

The body's process for tightly controlling blood sugar is well suited to analysis and modeling techniques devel-oped for feedback control systems. As illustrated in Figure 3, in a feedback control system, the value of the output parameter being controlled (blood glucose) is measured and fed back into the controller. This signal, along with other inputs, is continuously monitored by the controller, which adjusts the available system parameters to achieve the desired output.

System Diagram of the Metabolic Fuel Model

Figure 4 is a block diagram of the metabolic fuel model developed at Lincoln Laboratory. It incorporates dietary inputs, glucose level–sensing functions, glucose control signals, actuators, and the interconnectivity between

these elements in a nonlinear feedback control system. For the implementation described in this article, only fats and carbohydrates are considered. Body fat has a role as an energy storage mechanism and, in the model, all consumed fat is initially placed into a storage reservoir with unlimited capacity. In Figure 4, the movement and storage of fat are depicted in red. For digested carbohydrates, storage options are limited and are depicted in yellow, flowing into and out of the bloodstream in the form of glucose. The pancreas monitors the level of circulating glucose and secretes insulin to signal how to dispose of the excess glucose in the bloodstream. Each actuator is represented as a valve that responds to the amplitude of the insulin control signal. The metabolic rate is the sum of valves 2 and 4; the glycogen stores are controlled by valve 1; the macronutrient oxidation ratio is the ratio of valves 2 and 4; and converting glucose to fat, or de novo lipogenesis (DNL), is valve 3.

Also shown in this model is the concept of glucose overflow. When blood glucose becomes extremely high, additional mechanisms manifest in response, such as the kidneys passing excess glucose into the urine and other body tissues accumulating glucose. These responses are not considered actuation mechanisms in the model because they represent disease states.

THE INPUTS: FOOD AND ACTIVITY LEVEL

Food consumption, respiratory gas exchange (breathing), and energy demands (physical and mental activity) represent the body's metabolic interface to the outside world. Food intake, along with existing energy stores, provides the fuel needed to sustain metabolic processes. As much as 70 percent of the metabolic energy required over the course of a day is devoted to sustaining autonomic processes such as breathing, blood circulation, tissue repair and maintenance, digestion, and cognition [26].
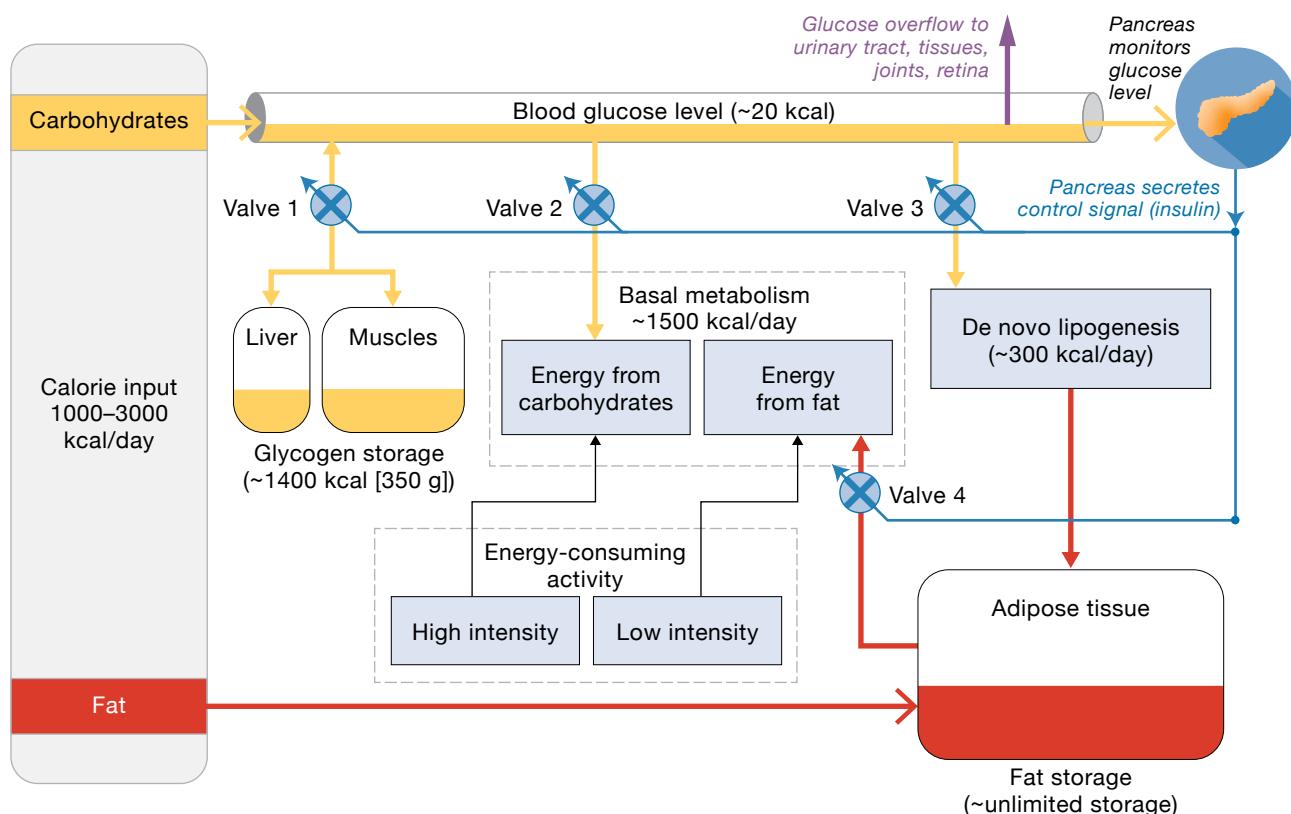
FIGURE 4. The metabolic fuel model features a system-level block diagram of the body's blood glucose feedback control system. The model includes dietary inputs, glucose level–sensing functions, glucose control signals, actuators, and the interaction of all these elements to regulate blood glucose level. Fat is depicted in red, while carbohydrates are depicted in yellow. The arrows represent the path of the calories in an individual, starting with digestion and moving into the body for storage or use.

Individuals also require energy to move about, maintain thermal regulation, and perform daily activities.

Ingesting certain foods, such as simple sugars, may rapidly introduce large quantities of glucose into the bloodstream and place tremendous demands on the blood glucose control system. In contrast, intense exercise may quickly deplete stored and circulating blood glucose, which must be replenished to sustain critical processes such as brain function. In both cases, the key is the rate at which the macronutrients enter and exit the bloodstream.

Food

Fat, carbohydrates, and protein are the three primary macronutrients, and each influences the dynamics of blood glucose circulation differently. For instance, fat does not directly influence the blood glucose concentration, as its digestion does not directly result in the production of glucose. In contrast, highly processed carbohydrates enter the bloodstream primarily as glucose, to be metabolized for energy, stored as glycogen, or converted into fat. Proteins can be converted into glucose through a slower and less efficient process termed gluconeogenesis [27] and are not included in the model of Figure 4.

Carbohydrates have the largest impact on circulating blood sugar and are thus a primary focus of our model. The rate of glucose's appearance in, and disappearance from, blood is critically important. The appearance of glucose in circulation depends on the timing of food consumption, quantity of carbohydrate consumed, rate of carbohydrate digestion and appearance in the blood stream, type of carbohydrate, and energy demands. The glycemic index (GI) is used as a measure of this parameter. Strictly speaking, GI is a measure of the speed with which food is turned into circulating blood glucose and is referenced to raw sugar, which is assigned a glycemic index of 100 [28]. The importance of glycemic response was demonstrated in a recent study in which the blood glucose response to various foods was tracked to establish personalized nutrition plans [29]. Our model allows a different GI to be set for each meal without requiring the GI to be specified for each food. With the addition of blood glucose monitoring, either directly or through an indirect measure such as the respiratory exchange ratio, the model can be tailored to particular individuals and foods. The respiratory exchange ratio is the ratio of the volume of carbon dioxide (CO_2)

produced to oxygen (O_2) consumed as measured in the exhaled breath of an individual and indicates the macronutrient mix supplying metabolic energy.

Metabolic energy generation at the cellular level is extremely complex. However, the overall impact of the multistep metabolic processes can be summarized by two relatively simple stoichiometric chemical equations that account for the O_2 consumed and byproducts (CO_2, water [H_2O], adenosine triphosphate [ATP]/energy) produced in the mitochondria per molecule of fuel oxidized.

CARBOHYDRATE BURNING (GLUCOSE)

$$C_6H_{12}O_6 + 6O_2$$
$$\Downarrow$$
$$6CO_2 + 6H_2O + 38ATP(energy)$$
$$RQ = moles\, CO_2 / moles\, O_2 = 6/6 = 1 \tag{1}$$
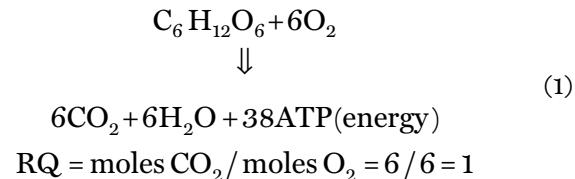
FAT BURNING (PALMITIC ACID) [30]

$$C_{18}H_{36}O_2 + 26O_2$$
$$\Downarrow$$
$$18CO_2 + 18H_2O + 131\,ATP(energy)$$
$$RQ = moles\, CO_2 / moles\, O_2 = 18/26 = 0.7 \tag{2}$$
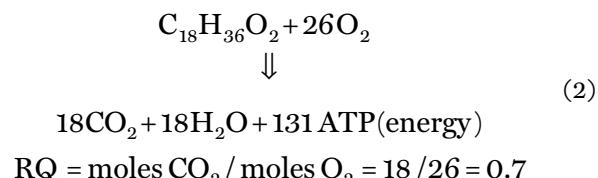
Equation (1) reveals that for every mole of oxygen consumed in the carbohydrate reaction, one mole of carbon dioxide is produced. This 1:1 stoichiometric relationship holds for all carbohydrates represented by $(CH_2O)_n$. For palmitic acid (a typical saturated fatty acid), 18 moles of CO_2 are produced for every 26 moles of O_2 consumed. Thus, the stoichiometry for fat burning (18/26 = 0.7) differs from that for carbs (6/6 = 1). The 0.7 ratio for palmitic acid is typical for many free fatty acids [31]. Therefore, by measuring moles of O_2 consumed and CO_2 produced, the relative fraction of carbohydrates versus fats that the body uses to meet metabolic energy needs is estimated.

The molar ratio between CO_2 produced versus O_2 consumed is called the respiratory quotient, or RQ, and is widely accepted as a key noninvasive indicator of metabolic processes occurring at the cellular level. If all of the oxygen consumed supports carbohydrate oxidation, then RQ is equal to 1, whereas if all the consumed

O_2 supports fat oxidation, the RQ is ~0.7. This simple concept can be extrapolated to an entire organism for which all the CO_2 exhaled and all the O_2 absorbed is measured and ratioed on a breath-by-breath basis to provide a reasonably good approximation of the sum of all the internal cellular chemical reactions. This system-level breath measurement is termed the respiratory exchange ratio, or RER. The RER is not identical to the RQ, but in practice, RER is often very similar to RQ, and the term RQ is frequently used interchangeably with RER [32].

For typical American diets, the glucose load into the bloodstream is driven by the digestion of carbohydrates. For instance, drinking a sugary beverage might deliver 50 grams of glucose (200 kcal) into the bloodstream within a fraction of an hour of consumption. If the glucose level is not offset in some way (see actuator discussion in later section), blood sugar levels could rise to over 1000 mg/dL (10 times normal) from a single drink. The rate at which high-GI carbohydrates release glucose into the bloodstream establishes a minimum necessary response for the metabolic control system to mitigate high glucose levels.

Activity Level

Understanding the timing, duration, and intensity of physical activity, whether it is part of daily life or formal exercise, is essential to correctly model blood glucose levels. The energy demands that physical activity places on the body, both in terms of intensity and duration, impact the rate of glucose depletion [33]. Low-intensity exercise is primarily fat burning and will have a negligible effect on glucose control. However, high-intensity exercise, which preferentially metabolizes glucose, affects blood glucose turnover and replenishment.

Strenuous exercise can raise metabolic demand to rates greater than 1000 kcal/hour, and much of this energy need is met by the utilization of glucose in neighboring muscle tissue. Typically, this glucose is supplied directly from internal glycogen stores and dietary intake, with both processes having important impacts on the functioning of the control system. High-intensity exercise sustained over a long period of time (e.g., running a marathon) can deplete glycogen stores and severely degrade performance in a phenomenon commonly known as hitting the wall or bonking.

THE SENSING FUNCTION

As an individual eats or exercises, the blood glucose–sensing function of the body detects changes in the circulating blood glucose level and will respond by signaling the body to act in a way to offset departures from nominal. Understanding the details of how insulin and its counterpart, glucagon, broadcast their control signals throughout the body is not essential to correctly model the functional behavior of the control loop. All that is required to implement a control model is to understand that subsystems in the body respond to increasing and decreasing blood glucose levels by adjusting the level of insulin and glucagon control signals, and that there are finite cellular response times associated with the generation of and response to the control signals.

THE ACTUATORS

The mechanisms by which the body responds to glucose control signals are called actuators. The term actuator is adopted from control theory terminology rather than medical terminology. For each actuator, it is important to consider how much capacity it has to offset the rates of glucose's introduction to, and removal from, the bloodstream. Control signals that exceed the actuator capacity limits introduce a nonlinear element to the system behavior and produce complex dynamic responses despite the relatively small number of functional elements in the model. Nonlinear effects are critical to understanding the relationship between type 2 diabetes, obesity, and metabolism. Table 1 lists the four actuators and provides nominal values for the rates of glucose control that they can achieve and the nominal limits (capacity) of their control authority. For each of these four actuators, it is noteworthy that there is a direct connection between the presence of insulin and the rise in the level of actuation, although knowledge of insulin's role in affecting glucose control is not required for understanding the model. Individual variance in the behavior of each of these actuators may change with time, aging, disease, or recent food intake and exercise.

METABOLIC RATE

In response to a signal that blood glucose is elevated, the body can react by increasing the basal metabolic rate to consume more glucose from the bloodstream [34]. This increase in metabolism may be achieved through

Table 1. The Four Primary Actuators Available to Reduce High Blood Glucose Levels and Their Nominal Rate/Capacity Values

CONTROL MECHANISM (ACTUATOR)	ACTUATION METHOD	GLUCOSE CONTROL RATE	CAPACITY LIMIT
Metabolic rate	Burn glucose at a higher rate	Moderate (50 grams/hour)	Maximum metabolic rate (50 grams/hour) [35]
Glycogen stores	Move glucose out of blood	Very high (250 grams/hour) [36, 37]	Glycogen stores (~400 grams typical) [38]
Macronutrient oxidation ratio	Increase burning of glucose for fuel	High (150 grams/hour)	0% reliance on fat (0 grams/hour)
Glucose conversion to fat	Remove glucose by conversion	Moderate (30 grams/hour) [38]	De novo lipogenesis (DNL) rate in liver (30 grams/hour)

a combination of increased body temperature, greater body motion, and other metabolic energy demands. For the purposes of the model, it does not matter how the metabolic rate increase is achieved, only that it occurs and is subject to rate and capacity limits. The metabolic rate cannot be arbitrarily increased in response to the signal, so its control authority to remove glucose is limited to approximately 50 grams/hour (for a nominal basal rate of ~2500 kcal/day [35]).

GLYCOGEN STORES

The body has the ability to store and release glucose outside the bloodstream by converting glucose into the starch-like polysaccharide glycogen, which can be stored in the liver or in skeletal muscle throughout the body. As indicated in Table 1, glycogen storage and release can occur rapidly, with glucose assimilation rates as high as 250 grams/hour [36, 37] and expenditure rates higher than 150 grams/hour [36, 39]. The total glycogen storage capacity in an adult body depends in part on muscle mass and can be as high as 800 grams [38] but is typically about 400 grams [40]. Once these glycogen stores are full, they can no longer serve as a glucose sink. Likewise, when glycogen stores are fully depleted, they cannot release glucose to compensate for declining blood glucose levels. While some carbohydrates, such as fructose, are directly converted to glycogen [41], our simplified model digests and passes all carbohydrates into the bloodstream

first. Similarly, the processes of storage and retrieval of glycogen in the liver and muscle are treated identically in this model despite their physiological differences.

MACRONUTRIENT OXIDATION RATIO

While metabolizing fat does not change blood glucose levels, metabolizing glucose directly depletes blood glucose levels. Therefore, shifting the metabolic fuel mix to favor glucose rather than free fatty acids is a favorable control mechanism for reducing high blood glucose levels. Adjusting the fuel mix allows the body to rapidly alter the rates of glucose consumption. Of course, oxidation balance cannot drive the mix beyond the extremes of 0 percent or 100 percent. Furthermore, since the blood/brain barrier inhibits fat as a source of brain fuel, some circulating glucose is necessary to sustain brain function (even for the case of very low-carbohydrate ketogenic diets) [42]. Maintaining cognitive brain function is a primary reason why blood glucose control is a priority in the body's control mechanisms.

CONVERT GLUCOSE TO FAT

Lastly, in terms of glucose disposal mechanisms, the body has the capacity to convert glucose into fat by means of DNL [43], which removes glucose from the blood, thereby reducing excessively high blood glucose concentrations. There is some debate about the maximum achievable DNL rates [44] and the conditions under which DNL

occurs in humans; however, measurements indicate that rates of upwards of 500 grams/day are achievable [38]. The DNL process seems to activate as a last resort when the control authority of the three other actuators becomes saturated. There is no direct fat-to-glucose reaction pathway in humans (other than the process of carbohydrates being produced from the small amount of glycerin released during lipolysis).

Validation of the Metabolic Fuel Model

To provide evidence of the validity of the model, we simulated the outcome of a comprehensive 14-day carbohydrate overfeeding study [38]. In the original study, three healthy young men, ages 21 and 22, weighing 62 to 72 kilograms (136.7 to 158.7 pounds), of height 174 to 180 centimeters (5 feet 9 inches to 5 feet 11 inches), and body fat 11 to 14 percent with no family history of diabetes or obesity, participated. The experiment lasted 14 consecutive days. During the first three days, the subjects consumed a high-fat low-energy (HFLE) diet, high in fat and low in carbohydrates, and followed an exercise program. The purpose was to deplete the glycogen stores of the individuals prior to tracking the energy balance over a period of 10 days. Halfway through this energy-restrictive period, the subjects were admitted into a whole room indirect calorimetry chamber, effectively a metabolic

monitoring chamber, in which respiratory exchange measurements and controlled feeding were continued for 10 days. After 36 hours in the chamber, the subjects' diet was changed to a high-carbohydrate low-fat diet for the following seven days. During the last two days, while still in the chamber, the subjects received a protein-sparing modified fast (PSMF) diet to begin returning their weight to normal. The subjects then left the respiration chamber but continued to consume the high-fat, low-energy, low-carbohydrate diet for two more days.

Figure 5 shows a dataset from this study that will be interpreted within the context of our model. As shown in Figure 5a, the subjects entered the metabolic chamber with glycogen stores depleted on day 3 and began consuming a high-carbohydrate diet. Initially, excess carbohydrates, beyond those required to meet metabolic energy needs, were converted to glycogen and stored in the liver and muscles. However, by day 5, glycogen stores were nearing their capacity and were unable to absorb all the excess dietary carbohydrates, at which point DNL activated to convert and store excess carbohydrates as fat. Figure 5b shows the corresponding change in the subjects' RQ that indicates the change in metabolic fuel substrate, with a resting RQ = 0.7, implying all fat oxidation; RQ = 1 implying all carbohydrate oxidation; and RQ > 1 implying DNL activity. Figure 5 shows increasing reliance on DNL
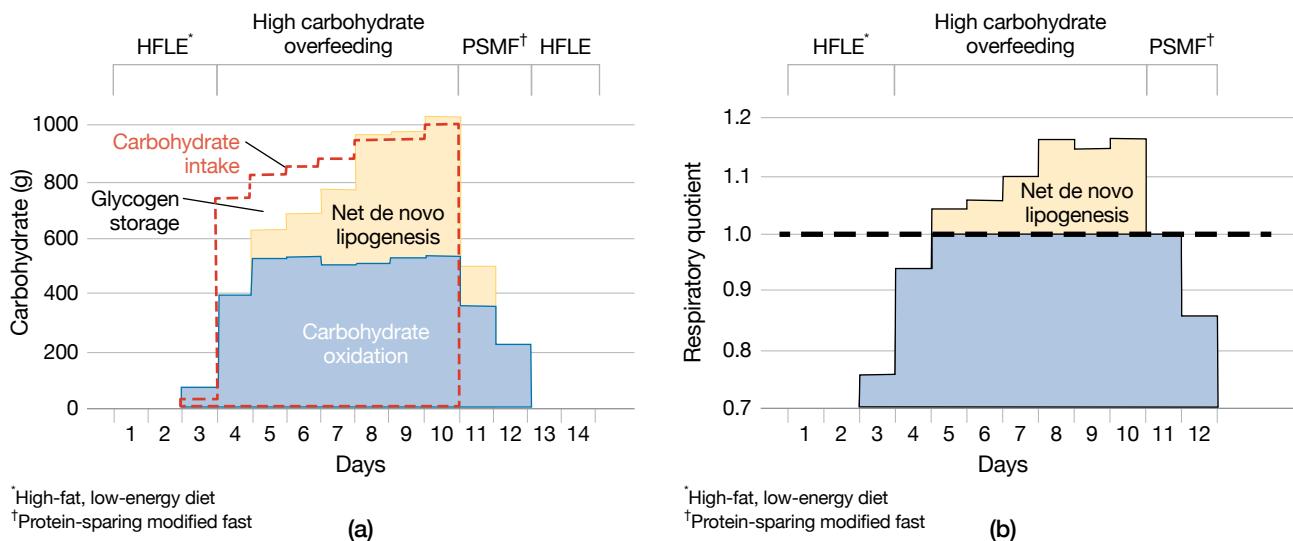
FIGURE 5. A carbohydrate overfeeding study was conducted to evaluate the effects of carbohydrate overfeeding, which led to the filling of glycogen stores and increasing liposynthesis (a), and to determine the respiratory quotient associated with the initial state of glycogen depletion, glycogen saturation, and de novo lipogenesis [38]. These results exemplify the utilization and saturation of the control valves in the theory. The results of this study are compared to the simulation results of the metabolic model.

to manage excess carbohydrates, with as much as ~500 grams/day of carbohydrate disposed via DNL, resulting in an increase of fat stores of about 2.5 pounds over the course of the seven-day low-fat diet (see Figure 6).

An executable version of the metabolic fuel model of Figure 5 was initialized with the data corresponding to the average dietary intake, activity, and physiology of the three subjects in the carbohydrate overfeeding experiment. Unlike the original study, which took 14 days to complete, the simulation of the study was executed in less than one second.

Four of the measured parameters from the study are compared to the simulation output in Figure 6. The solid curves represent the daily average metabolic values measured for the three individuals involved in the study. The dotted curves represent the values predicted by the simulation model given the feeding and exercise schedules followed during the study. The largest disparity between the simulation and the actual study results is in the predicted value of RQ during the days in which DNL was active. The disparity is due in part to the fact that the RQ associated with DNL is dependent on the type of fat synthesized and can vary from 2.75 to 9. For an example, see Equation (4) on page 128. A significant point of agreement between the study data and the model-based simulation data is that the simulated RQ rose above 1 on day 5, predicting the onset of DNL on the day it was observed in the actual experiment.

The relatively close agreement between the published data and the simulated output provides compelling evidence that, without resorting to detailed modeling of cellular metabolic pathways and processes, the impact of dietary macronutrients and exercise on metabolism can be captured by this system-level feedback control model.
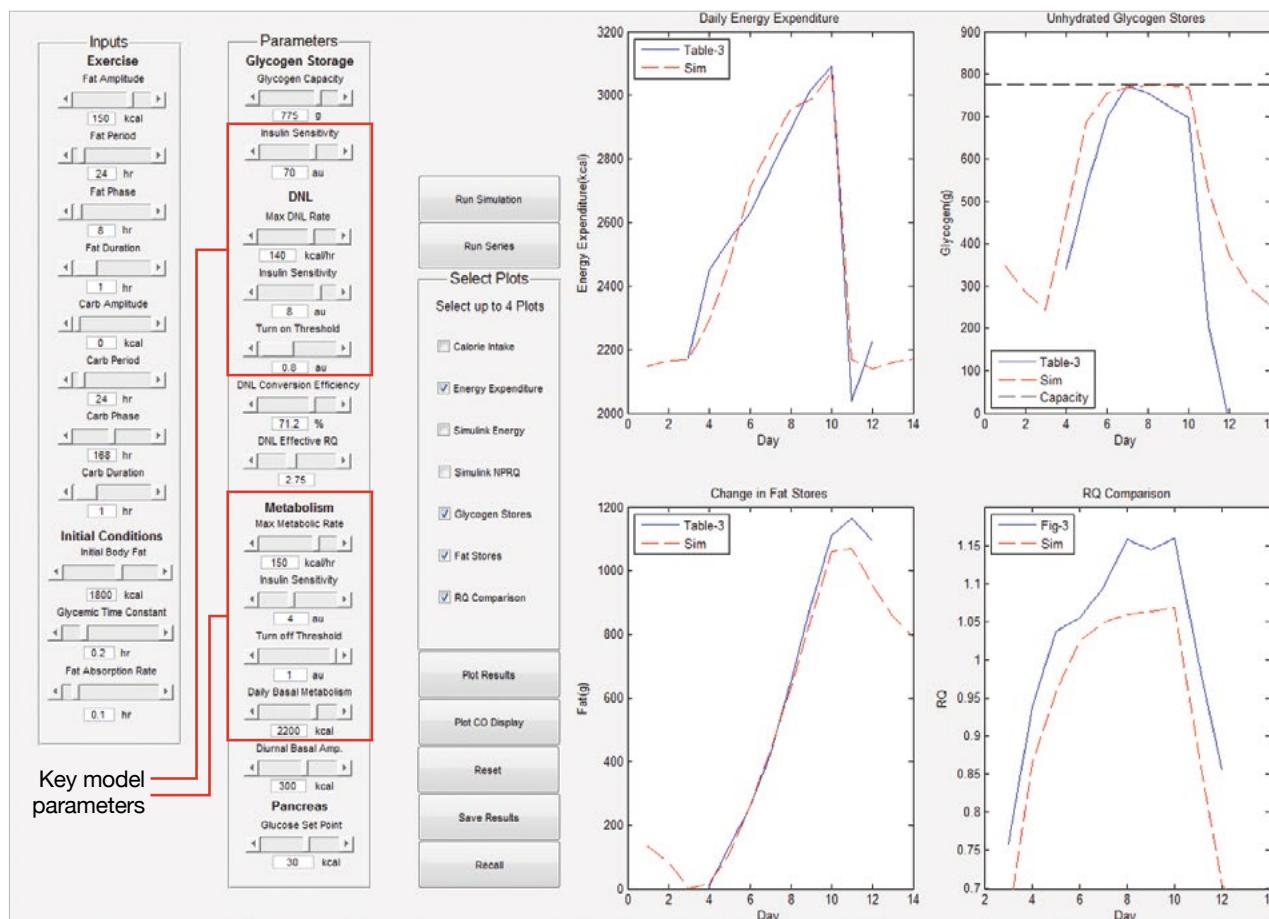
FIGURE 6. The metabolic model simulation inputs and prediction performance showed a high agreement with the results of the published overfeeding study, indicating that the metabolic model is a promising system for predicting the real-life impact of diet and exercise on metabolism. The key tunable parameters of the model are highlighted in the red boxes.

Implications of the Metabolic Fuel Model

Applying control systems theory to model human metabolism provides a powerful simulation tool for predicting the metabolic response (blood glucose level control, fuel usage, glycogen storage, fat storage, and insulin level control) that an individual will exhibit over a wide range of diets and activity profiles.

The realization that blood glucose control is a critical function that takes precedence over other functions of the homeostatic control system, such as body weight stabilization and energy balance, can lead to a view of the causes and treatments for excess body weight and type 2 diabetes that is different from the view typically encountered. Table 2 compares and contrasts some of the implications drawn from the metabolic fuel model to commonly held views [45].

Utility of the Metabolic Fuel Model

To illustrate one of many possible applications of the metabolic fuel model, we simulated a diet experiment. The simulation provided insight into why a dietary model that only considers calories, and not the macronutrient composition of those calories, is inadequate for understanding and managing body weight.

For one week, the simulated subject was fed a high-carbohydrate diet with a caloric ratio of 80 percent carbohydrates to 20 percent fats. A glucose-tolerance test, consisting of an injection of 100 kcal of glucose (equivalent to drinking about eight ounces of soda) over a 10-minute span, was administered to the simulated test subject. After injection of the glucose into the bloodstream, the four actuators, shown in the block diagram in Figure 4, were monitored as the glucose spike induced by the sugary soda was cleared to restore the baseline blood glucose concentration.

In the second week, the same simulated subject was fed a diet consisting of the same number of calories as the first week, but with a low carbohydrate ratio (20 percent carbohydrates and 80 percent fats). The same glucose injection was repeated. The glucose recovery results corresponding to the two feeding regimens are summarized in Figure 7, and the results show markedly different responses to the same glucose challenge.

- The glucose clearance time is 50 percent longer following the high-carbohydrate/low-fat diet.
- There is evidence of what has been termed insulin resistance in the high-carbohydrate diet (it takes three

Table 2. Implications about Various Medical Conditions Drawn from the Metabolic Fuel Model versus Commonly Encountered Views about the Same Conditions

METRIC/RESPONSE	CONVENTIONAL VIEW	MODEL VIEW
Obesity	Contributor to type 2 diabetes [46, 47]	Two of four mechanisms that the body employs to avoid becoming diabetic lead to obesity
High insulin levels	Caused by insensitivity to insulin (insulin resistance)	Root cause is chronically overwhelming the four glucose control processes
Diet for weight loss	Requires overt control of energy balance (calorie intake/expenditure)	Requires active avoidance of macronutrient imbalance (carbohydrate overconsumption)
Exercise for weight loss	Low-intensity exercise is best because it is fat burning	Choose exercise to counteract macronutrient imbalance (e.g., high-intensity exercise to counteract carbohydrate overconsumption)
Feedback for guiding weight loss intervention	On-demand body weight measurements (scale)	On-demand metabolic state measurements (e.g., RQ, insulin, etc.)
Excess body weight trend in the population	Root causes are inactivity and overeating	Root causes are historical changes in dietary macronutrient mix that subvert homeostatic control loops

times more insulin to clear the glucose after consuming the high-carbohydrate/low-fat diet).

- The DNL process activated for the high-carbohydrate/low-fat diet to convert and store excess glucose as fat.
- There was a reduced capacity for converting glucose into glycogen following the high-carbohydrate/low-fat diet.
- There was a reduced opportunity to dispose of excess blood glucose by means of fat burning on the high-carbohydrate/low-fat diet.
- There was a prolonged and exacerbated period of low blood sugar following glucose clearance on the high-carbohydrate/low-fat diet.

The results in Figure 7 also suggest how the promotion of low-fat (consequently, high-carbohydrate ratio) diets over the past 40+ years may have contributed to the dramatic rise in obesity. As shown by the simulation, even small amounts of dietary fat may be stored to reduce high blood sugar levels by preferentially using glucose to meet energy demands (resulting in fat storage in adipocytes). In the case of high-carbohydrate diets, DNL is also more likely to be activated to convert the excess carbohydrates into stored fat.

Diabetes and the System Model

A principal characteristic of diabetes is an impaired ability to properly control blood glucose concentration. Because the foundation of the metabolic fuel model presented here is blood glucose control, the model has direct applicability to understanding the underlying causes of diabetic behavior. In the case of type 1 diabetes, from the metabolic fuel model perspective, the impairment is in the sensing system. For type 1 diabetics, despite an elevation in blood glucose, the feedback signal is very weak or nonexistent and not fed to the four actuation mechanisms. Without the ability to produce insulin and broadcast the control message, the actuators are never enabled despite having ample capacity to control blood sugar. When insulin is injected into the bloodstream, the full set of actuation mechanisms are activated to control blood glucose levels. Maintaining the precise level of insulin required to balance incoming glucose is very difficult to achieve and requires constant monitoring of blood glucose level. Understanding the feedback mechanisms may help type 1 diabetics better manage their blood sugar in response to different dietary and exercise inputs.

(a)

(b)
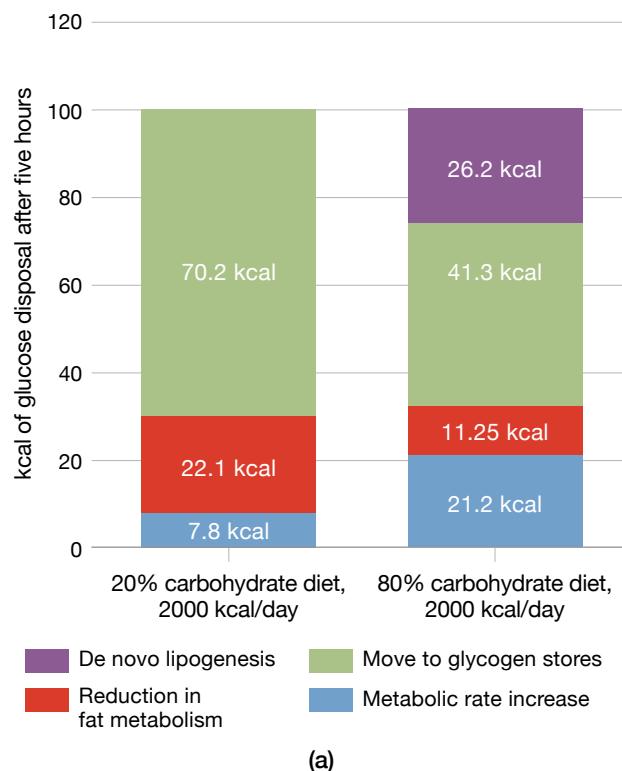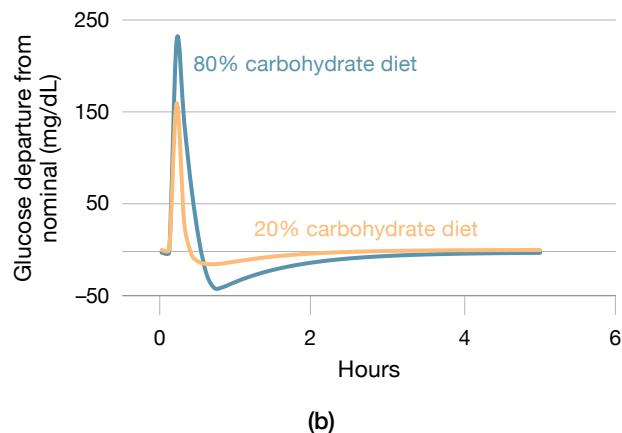
FIGURE 7. This figure shows the results of simulated glucose infusions following low-carbohydrate and high-carbohydrate diets. The four colors correspond to the four glucose disposal mechanisms discussed in Table 1 (a). For the same calories during the week preceding the glucose infusion, the low-fat diet resulted in a less desirable metabolic response, including fat gain through de novo lipogenesis, and higher elevated glucose levels followed by a longer low blood sugar event (hypoglycemia) (b).

While the metabolic fuel model does not provide any new insight into the cause of type 1 diabetes, the model is particularly useful in providing insights into the management and avoidance of type 2 diabetes. In type 2 diabetes, the body is also unable to adequately control blood glucose levels. However, unlike type 1 diabetes, type 2 diabetes results not from a malfunction of the insulin signal generation system but rather from a problem with the control mechanisms. Typical insulin levels with type 2 diabetes are much higher than normal, yet the blood glucose control actuators do not respond strongly enough to the elevated insulin signal to effectively control the glucose levels. This condition is referred to as insulin resistance and suggests that the actuators need higher signal levels than previously required to achieve proper blood glucose control. However, there is no universal agreement or consensus regarding the root cause of this apparent loss in insulin sensitivity.

The metabolic fuel model offers a coherent explanation of the precursors to, and root cause of, the symptoms of type 2 diabetes. If all of the four actuation mechanisms incorporated in the model are functioning at their full capacity, then the presence of higher levels of insulin will have a minimal effect on the rate at which glucose is cleared from the blood. Consequently, symptoms of type 2 diabeties will be observed (insulin resistance), and there will be high levels of both glucose and insulin. Blood glucose control fails when the control authority available in the actuators is less than the rate of glucose appearance in the bloodstream or, put another way, if too much sugar is in the blood and no actuator capacity is left to remove it. This interpretation of type 2 diabetes implicates the macronutrient content ingested, as opposed to a degradation of the glucose processing mechanisms, as the root cause of diabetic symptoms. Any individual whose actuation mechanisms cannot keep up with the rate at which carbohydrates are being absorbed will exhibit diabetic symptoms. Changes to the available food supply that increase consumption of carbohydrates and high-GI foods challenge the available control authority of a great fraction of the population. The metabolic fuel model explanation for the origins of type 2 diabetes does not require postulating or explaining the appearance and mechanism of insulin resistance, nor does it preclude the development of insulin resistance over time in response to chronic carbohydrate overloading.

Conversely, the metabolic fuel model suggests that if carbohydrate consumption is reduced to a level compatible with the individual's control authority, the symptoms of type 2 diabetes will quickly disappear. There is ample evidence in the literature that carbohydrate reduction can successfully reverse the progression of type 2 diabetes [5, 48–50]; however, the science is not settled [51].

Obesity and the Metabolic Fuel Model

The motivation for creating the metabolic fuel model was to functionally express the mechanisms involved in the body's response to changes in blood glucose concentration. However, the insights provided by the metabolic fuel model extend well beyond blood glucose control, including how these blood glucose control mechanisms impact long-term changes in body weight.

Contradicting the view that obesity is the primary cause of insulin resistance and the onset of type 2 diabetes [45, 46, 52], the metabolic fuel model suggests an alternative sequence of events. The model reveals that when the body is struggling to reduce the blood glucose level, it will stop burning fat and, in some situations, convert excess blood glucose into fat (DNL) in an attempt to keep up with the rate of glucose input. If, through dietary choices, an individual is constantly releasing high levels of glucose into the bloodstream, one predictable consequence of these control mechanisms, as summarized in rows three and four of Table 1, is an increase in stored fat. The metabolic fuel model further suggests that for an overweight individual seeking to reduce body weight through a low-fat high-carbohydrate diet, the metabolic response of the body to control circulating blood glucose acts in opposition to the desired fat burning, resulting in weight gain rather than weight loss.

The metabolic fuel model can also be used to predict an individual's response to a given diet and, in particular, how the response varies with the number of calories, the macronutrient mix, the GI, and the duration and intensity of exercise. Tuned to an individual, the model can be used to determine what dietary intake and nutrient mix will ensure that the blood glucose control does not override the body's hormonal signals related to weight control, enabling the individual to preferentially burn dietary and stored fat rather than carbohydrates to meet metabolic energy needs.

Metabolic Sensing

The metabolic fuel model provides a framework for understanding the impact of macronutrient imbalances on weight management, obesity, and type 2 diabetes, and also provides a means of predicting the impact of diet and exercise on an individual's metabolic health and performance. To apply the model to a particular individual or class of metabolically similar individuals, standard physiological data such as age, gender, weight, and height can be used to estimate basal metabolic rate, glycogen stores, and responses to exercise. However, to tune the model to a particular individual or class of individuals, such as athletes or military trainees, measurement of the individuals' metabolic responses to diet and exercise is required.

Metabolic Fuel and Energy Measurement Confounds

While the stoichiometric relationships described by Equations (1) and (2) are relatively straightforward, proper interpretation of the measured values of O_2 consumption and CO_2 production requires awareness of a number of temporal and physiological confounds. For example, CO_2/bicarbonate buffering in the bloodstream can lead to previously buffered CO_2 being released during hyperventilation, and CO_2 is also released during the conversion of glucose to fatty acids and anaerobic energy production. Each of these confounds can result in RER measurements greater than 1, so the equivalence between cell-level metabolism (RQ) and the molar ratio of breath gases (RER) needs to be applied with caution [31, 32, 53]. Alcohol is another confound since it cannot be stored and its disposal is a priority. So, as with carbohydrate overconsumption, alcohol may further suppress the metabolism of dietary fats. However, unlike that of carbohydrates, the RQ of alcohol is less than 0.7, so the metabolization of alcohol pushes the RQ down rather than up.

Protein is not a preferred source of metabolic energy but, when present in excess, the amino acids are processed and stored as glucose or ketones. The nitrogen waste that is liberated in the process is converted to urea in the urea acid cycle and eliminated through the urine. The RQ for proteins is nominally 0.85, midway between the RQ for fats and carbohydrates. Consequently, while protein can serve as a source of metabolic energy, the bias it introduces when neglected in the estimate of fat versus carbohydrate oxidation is typically a small portion of total calories consumed. If necessary, the average contribution of protein to metabolic energy can be ascertained by collecting and analyzing the nitrogen content of the subject's urine.

Equations (1) and (2) not only establish which type of fuel is being used, but they also enable estimation of the amount of energy created in each reaction. Given the RQ and the volume of O_2 consumed, the energy expenditure (EE) can be estimated from the Weir equation [54], where EE is the total energy expenditure and VO_2 is the time-dependent volumetric rate of oxygen consumption.

$$(3)$$

$$\text{EE}\left(\frac{\text{kcal}}{\text{day}}\right) = \left[3.9 + 1.1 \times \text{RQ}\right] \times 1.44 \times \text{VO}_2\left(\frac{\text{mL}}{\text{min}}\right)$$

Note that because RQ is normally between 0.7 (all fat) and 1 (all carbohydrate), the accuracy of EE calculated from the Weir equation is dominated by the accuracy of the VO_2 measurement. A completely erroneous estimate of RQ (e.g., an RQ estimate of 1 when the true value is 0.7) results in less than a 7 percent error in the estimate of expended energy, provided the VO_2 value is correct. However, errors in RQ and VO_2 may be correlated, resulting in a compounding effect on the error in the energy estimate.

The EE and RQ are both useful quantities for the metabolic fuel model. The measurements required to compute RQ and EE enable the calculation of additional parameters of interest, such as the VO_2, respiration rate, breath volume, and metabolic equivalent of task (MET).

RQ as an Indicator of Macronutrient Mix

Figure 8 shows the nominal range of RQ levels and its relationship to metabolism for measurements of physically active and resting subjects. When the body is inactive, metabolic demand is mostly driven by breathing, blood circulation, body temperature control, and other autonomic processes. As the voluntary activity level increases, these autonomic metabolic demands are overshadowed by muscle demand, and the basal metabolic energy needs are overshadowed by the substrate balance needed to meet the proportionally greater muscle energy demand. According to Equations (1) and (2), an RQ of 0.7 implies that all of the calories are supplied through

fat combustion, whereas an RQ = 1 implies that calories are being supplied exclusively from glucose. Figure 8 implies that resting RQ is driven primarily by the available mix of circulating macronutrients, whereas the preferred fuel (fats versus carbohydrates) during exercise is driven primarily by exercise intensity [55]. When the RQ is between 0.7 and 1, the metabolic fuel substrate is a combination of carbohydrates and fats. A resting RQ greater than 1 indicates that DNL (conversion of glucose into triglycerides) is occurring. The RQ associated with DNL depends upon the specific form of fat conversion. For example, the conversion of glucose to a fat (palmitoyl-stearoly-oleoyl-glycerol) is given by the stoichiometric chemical equation
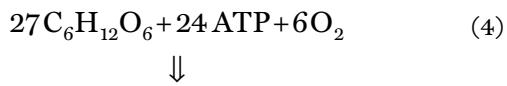
$$27\,C_6H_{12}O_6 + 24\,ATP + 6\,O_2 \qquad (4)$$
$$\Downarrow$$
$$2\,C_{55}H_{104}O_6 + 24\,ADP + 52\,CO_2$$
$$RQ = moles\,CO_2\,/\,moles\,O_2 = 52\,/\,6 = 8.7$$

The RQ of 8.7, when averaged with the total RQ across all cells, results in a whole body RQ >1.

A resting RQ greater than 1 also implies that fat burning has largely ceased, and all of the metabolic fuel is coming from carbohydrates. Alternately, an RQ above 1 produced during strenuous activity indicates that the oxidation needs of muscle exceed the oxygen supply and that anaerobic reactions are occurring. The anaerobic glycolysis does not require an O_2 input to the stoichiometry and temporarily skews the RQ calculations to higher values [56]. After exercise, the resulting pH drop from lactic acid production shifts the bicarbonate buffer system toward releasing stored CO_2 into the exhaled breath. Recognizing and understanding the physiology of confounds such as anaerobic glycolysis is important when one performs breath-based (i.e., RER or RQ) energy measurements that assume aerobic metabolism.

As shown in the metabolic fuel model in Figure 4, there is a relationship between RQ, EE, and the metabolic actuators. The state of three of the four actuator values can be inferred by measuring EE and RQ. If the resting RQ is above 1, indicating high blood glucose levels, the model predicts that valve 4 (fat burning) is completely closed and valves 2 (carb burning) and 3 (DNL) are open, allowing for only carbohydrate oxidation and conversion
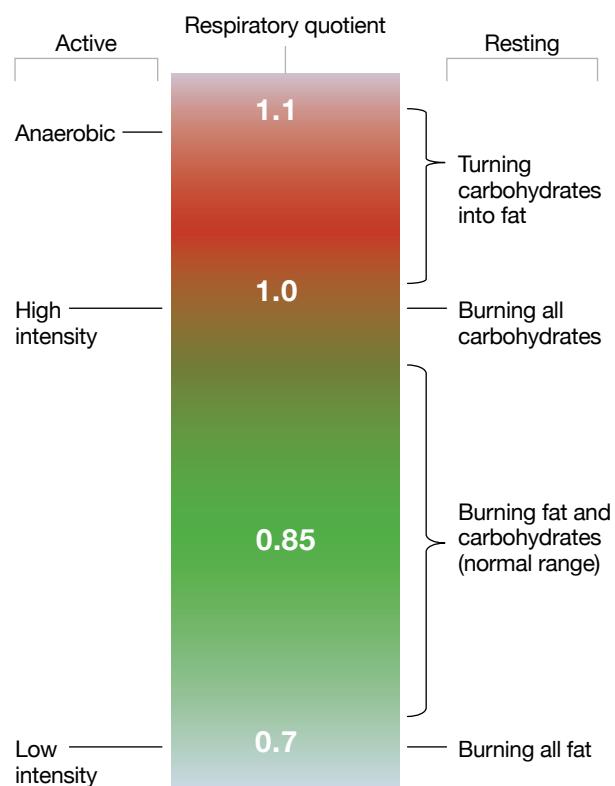
FIGURE 8. This figure shows the normal range of the respiratory quotient (RQ) and its significance for resting versus physically active measurements. Resting RQ is primarily driven by the available mix of circulating macronutrients, whereas the preferred fuel during exercise is driven primarily by the intensity of exercise.

into fat. When the homeostatic control system is operating normally, the resting RQ indicates the ratio of fat to carbohydrate oxidation, and therefore the RQ is isomorphic to knowing the flow relationship between valve 4 to valve 2. From Equation (3), the total number of calories metabolized is given by the sum of carbohydrate and fat oxidation. This energy flow is captured in valves 2 and 4. Therefore, calculations of RQ and EE unambiguously define the settings of valves 2, 3, and 4 in the metabolic fuel model, leaving only the available glycogen storage capacity uncertain.

Weight management protocols frequently focus on calorie consumption as the only metric of consequence. The model of Figure 4 shows that a more complete understanding of metabolic state and macronutrient imbalances can be achieved by measuring and tracking the dynamics of both RQ and EE.
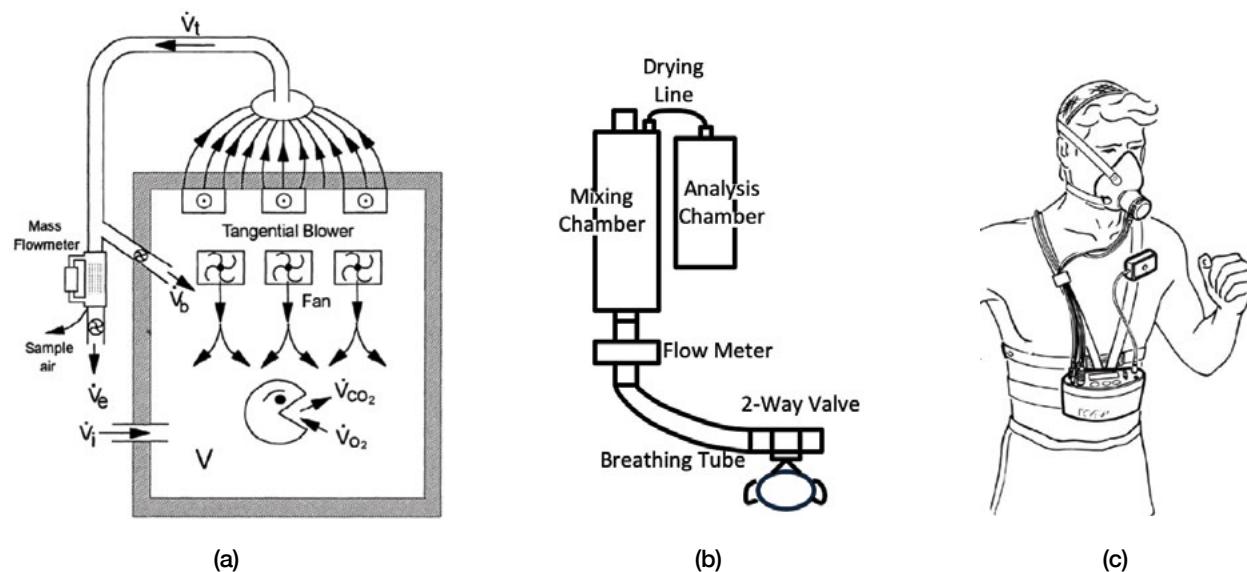
FIGURE 9. This illustration shows diagrams of the three different metabolic sensor classes, including a whole room calorimeter design (a) [57], a mixing chamber apparatus employed in single-user laboratory measurement systems (b) [58], and a breath-by-breath instrument designed to support mobile use (c) [59].

Traditional Methods of Measuring the Respiratory Quotient and Energy Expenditure

In general, three classes of systems are available for measuring RQ and EE. Each is described briefly in the following sections and represented schematically in Figure 9. Understanding the traditional methods of indirect calorimetry provides a context for understanding the distinctive features of a novel prototype sensor that significantly reduces the cost and complexity of traditional measurement methods.

WHOLE ROOM INDIRECT CALORIMETRY

Whole room indirect calorimetry involves isolating the subject in a small room, typically 10,000–20,000 liters in volume, and precisely measuring both the gas composition of fresh makeup air flowing into the room and the volume flow and gas composition of the exhaust air, as illustrated in Figure 9a [60]. The whole room indirect calorimetry approach allows for an unencumbered range of movement and activity within the room, which facilitates the study of exercise, sleep, and other specific activities over long intervals of time, typically on the order of days. Because a 20,000-liter whole room calorimeter corresponds to a volume of more than 20,000 breaths under resting conditions, it may take hours for new breaths to diffuse, mix, and produce a sufficiently

strong signal to be measured [57, 61]. Transient events with durations of less than a few hours are difficult to measure reliably, and the mass spectrometer needed to measure the very small changes in room air composition attributed to changes in consumed O_2 and exhaled CO_2 is quite expensive. More importantly, not all activities of interest to the military or athletes can be conducted in a 20,000-liter room.

MIXING CHAMBER IMPLEMENTATIONS (METABOLIC CARTS)

A mixing chamber (integrated breath sensor) employs a breath-mixing chamber with a nominal volume of three to four liters to capture one or more exhaled breaths from the subject even during periods of exercise involving large tidal (total) volumes. The chamber is designed so that as each new breath is exhaled, it mixes with the previously expired breaths while displacing a proportional amount of the existing breath sample mixture from the chamber. The mixed breath in the chamber is continuously sampled by a pump and sent through series-connected, high-speed (20 or more measurements per second) gas sensors to measure the changing concentrations of O_2 and CO_2. The measurements provide a moving average of the test subject's RQ and EE, with effective averaging time determined by the volume of the mixing chamber relative to the

tidal volume of each breath. Since the gas concentrations of each breath exhaled into the mixing chamber are not diluted by room air, the gas sensors in the metabolic cart implementation can be smaller, cheaper, and less sensitive than those required for whole room indirect calorimeters. Mixing chamber–based breath measurements have the advantage of providing a temporal resolution on the order of a couple breaths, and the entire measurement system can be placed on a mobile cart (or metabolic cart) and wheeled to treadmills or other activity areas to perform clinical or lab measurements on demand. The ownership and operating costs of a metabolic cart are significantly lower than those of a whole room indirect calorimeter. However, metabolic carts generally require wall plug power, and their interface and overall size and cost confine their use to research laboratories and clinics, making them impractical for field studies and too expensive for widespread personal ownership.

## BREATH-BY-BREATH IMPLEMENTATION (MOBILE SENSORS)

Portable sensors for field use generally employ a breath-by-breath measurement technique. In a breath-by-breath system, the user wears a mask capable of capturing and measuring the entire exhale (main stream) volumetric flow rate by means of a spirometer, while a pump continuously samples a fraction (side stream) of the respiratory flow and delivers the sampled gas to a series-connected pair of fast (25–50 samples/second) O_2 and CO_2 sensors that measure gas concentration many times during each breath cycle. The 25- to 50-hertz sampling rate is required because, for intense exercise, a complete inhale/exhale breath cycle can be as short as one second. Software then pairs the sample-by-sample measurements of volume flow rate with gas concentration measurements that have been corrected for the time delays associated with sequential measurement of the gas concentrations. The software computes a continuous series of differential volume elements that, when integrated, enable computation of volume flow rates for O_2 consumed and CO_2 exhaled as a function of time. Rapid, on-the-fly measurement of CO_2 and O_2 eliminates the requirement for a mixing chamber and enables metabolic estimation with a temporal resolution limited only by the time constants of the gas concentration sensors. Eliminating the mixing chamber produces a compact portable sensor that can be used outside a laboratory setting and addresses a larger range of use cases. However, the measurements of flow and gas composition must be performed rapidly, and for a given aliquot of gas, the flow, O_2, and CO_2 measurements all occur at slightly different times. Consequently, calibration of breath-by-breath systems requires precise time alignment of the instantaneous flow measurements with the sequential gas concentration measurements, each of which is made at a slightly later time. The need for time coherence among all three sensors may be difficult to maintain over a variety of environmental conditions and activity levels, and verifying calibration in the field is difficult without supporting calibration equipment. The high sample rate requires fast gas sensors that increase the system power consumption and cost. One of many applications for breath-by-breath mobile sensors is metabolic measurement of athletes as they train (some sensors have even been adapted for use by swimmers). However, the high cost and calibration complexity make breath-by-breath sensors unattractive for large-scale group studies or personal ownership.

Prototype Personal Metabolic Sensor

Both the Army and the Marines are interested in measuring RQ and EE in the field on a cohort of dozens of soldiers as they conduct a variety of load-bearing activities over the course of many hours or days. To make concurrent metabolic field measurements for dozens of soldiers affordable, the individual sensors must be low cost and easy to use without extensive training. The sensor must also be sufficiently small and lightweight to minimize impact on the normal functions performed by the soldiers. Low cost, ease of use, and small size are also attractive qualities for athletic use of the sensor in schools and sports programs and for personal ownership and metabolic health tracking.

Cost, size, weight, and power are often invoked as the primary metrics for comparing mobile systems that perform a similar function. However, because the power required by a mobile sensor over the period of operation translates into battery weight, the comparison can be reduced to just three parameters: cost, size, and weight (CSaW). Figure 10 compares the CSaW of Lincoln Laboratory's prototype sensor under development, named the Carbon-dioxide/Oxygen Breath and
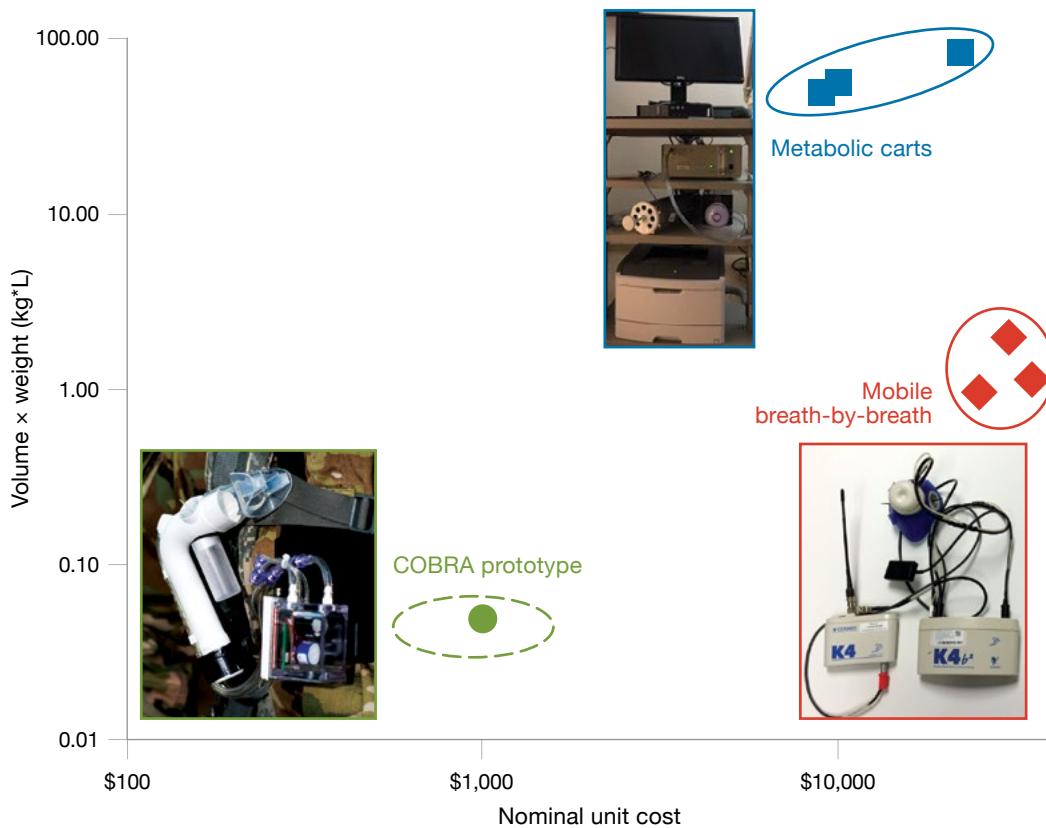
FIGURE 10. The prototype COBRA sensor has lower volume, weight, and estimated cost in comparison to commercially available metabolic carts and mobile metabolic sensors.

Respiration Analyzer (COBRA), to nominal values of CSaW associated with representative commercial metabolic carts and mobile sensors.

We achieved the significant reduction in CSaW of the prototype COBRA sensor by means of a novel passive proportional breath-sampling scheme [62–65] that eliminates the need for mechanical pumps or valves and replaces the conventional multiliter mixing chamber with a miniature mixing chamber that is about 100 times smaller. Relative to the breath-by-breath measurement technique, the passive proportional sampling technique enables the use of slower, less expensive gas sensors. Consequently, the total combined cost of the constituent O_2, CO_2, and flow sensors in single unit quantities is less than $250, and this cost is expected to be reduced in production with volume purchasing.

The prototype COBRA sensor system is shown in Figure 11. Figure 11b shows the COBRA sensor in use with a chest harness to enable hands-free operation. The mixing chamber currently employs a GoPro

bayonet mount. The chest harness is only one of many GoPro-compatible mounting options, including shoulder mounts, backpack strap mounts, and helmet mounts, that enable the users to select mounting options that are best suited to the units' intended use.

Passive Proportional Side Stream Sampling Innovation
As shown in Figure 11a, to record data with the COBRA sensor, the user simply begins breathing through the flow tube in which the inhale and exhale breaths pass through a specially designed internal geometry. Upon exhale, the flow tube creates a positive pressure difference between the entrance and exit ports of the flow tube. The positive pressure difference diverts a small fraction of exhaled breath into the mixing chamber in proportion to the volumetric flow rate of the exhaled breath. Additionally, the location of the high- and low-pressure sample ports on the flow tube creates a null pressure difference during inhale when gas is flowing in the opposite direction.

The result is that the flow tube acts as a diode, diverting exhaled breath into the mixing chamber for analysis while preventing ambient air from diluting the mixing chamber contents during inhale. A plot of the gas collection percent versus the average flow rate through the flow tube is shown in Figure 12. The tight clustering of the samples over a wide range of expiratory flow rates (i.e., over a wide range of exercise intensity) is evidence that the volume of gas diverted to the mixing chamber at any instant is proportional to the instantaneous exhale flow rate. As a result, the sample diverted to the mixing chamber preserves the constituent gas ratios in the exhale breath, and the sample is a proportional fraction of the entire exhale volume. Similarly, the tight clustering of samples during inhale and the relatively small proportionality constant indicate minimal dilution effects over a wide range of flow rates despite the absence of a valve on the mixing chamber.

This passive proportional sampling technique [62–65] enables more than a hundredfold reduction in the volume of the mixing chamber relative to that of whole breath–integrating metabolic carts while avoiding the need for fast gas sensing required in breath-by-breath systems. In addition, the time to collect and mix samples from several breaths is commensurate with the measurement time constants of low-cost miniature gas sensors. The passive proportional gas collection technique preserves the temporal relationship between
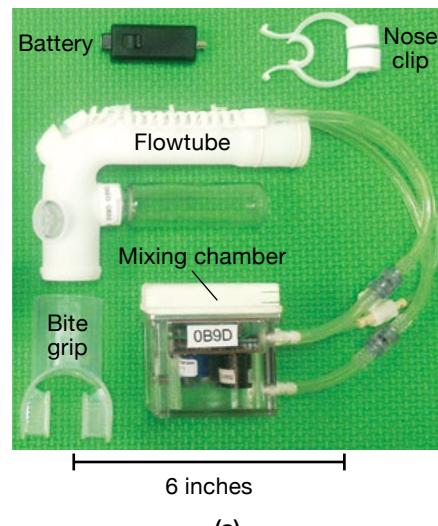
(a)
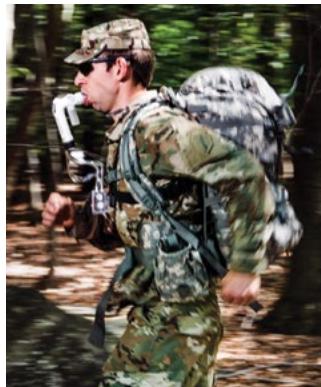
(b)

FIGURE 11. The lightweight and easy-to-use design of the COBRA sensor system includes a rechargeable battery, nose clip, and bite grip (a). The COBRA sensor is compatible with GoPro camera-mounting harnesses, enabling hands-free metabolic measurement during exercise (b).
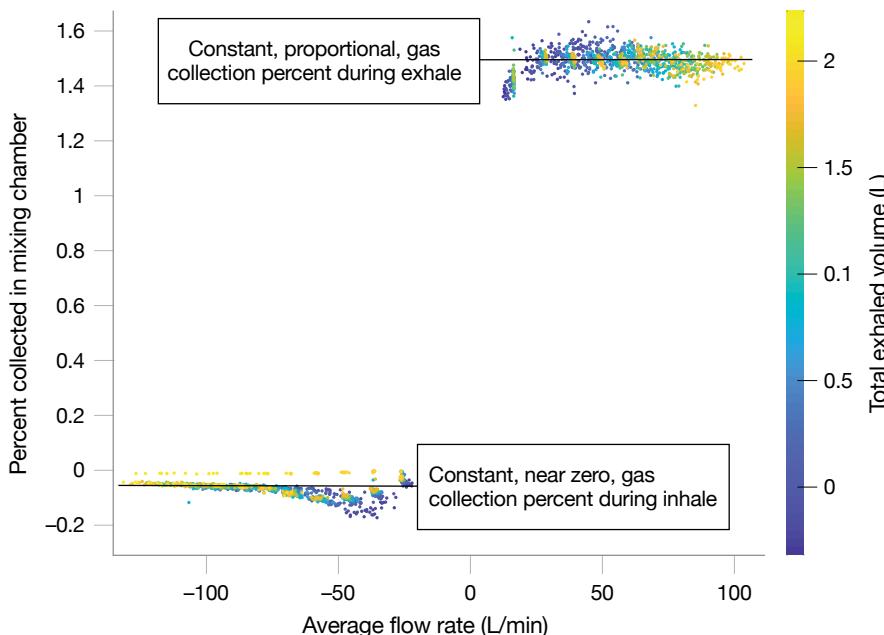
FIGURE 12. This plot reveals the proportional sampling and diode features of the Carbon-dioxide/Oxygen Breath and Respiration Analyzer flow tube design. These data show that the fraction of gas collected in the mixing chamber during an exhale is proportional to instantaneous flow rate over a wide range of flow rates, resulting in a relatively constant sampling fraction of 1.5 percent of the total exhale gas volume. Additionally, the negative flows, which represent dilution during inhale, are relatively small, typically less than 0.1 percent.

gas concentration and volumetric flow without requiring the high-speed gas sensors or large mixing chambers of existing commercial systems.

COBRA Data Products and Validation

In the COBRA sensor, flow and gas composition data are collected in real time and stored internally for later download. In addition, data can be processed by the sensor to produce derived data products and passed via a Bluetooth low-energy wireless connection to an Android device for real-time display. Data from metabolic measurements are presented in two different formats, either as a series of panel plots or as a Microsoft Excel–compatible spreadsheet. Data products include instantaneous flow, respiration rate, tidal volume, minute volume, VO_2, VCO_2, RQ, and EE. Three examples of the many available data products are shown in Figure 13.

The plots in Figure 13 correspond to an interrupted time series of three different levels of exercise intensity from one subject, with the walking data appearing first, followed by the jogging data and finally the running data. Each panel shows a different data product, with the EE in the top panel, RQ in the second, and the volumetric flow
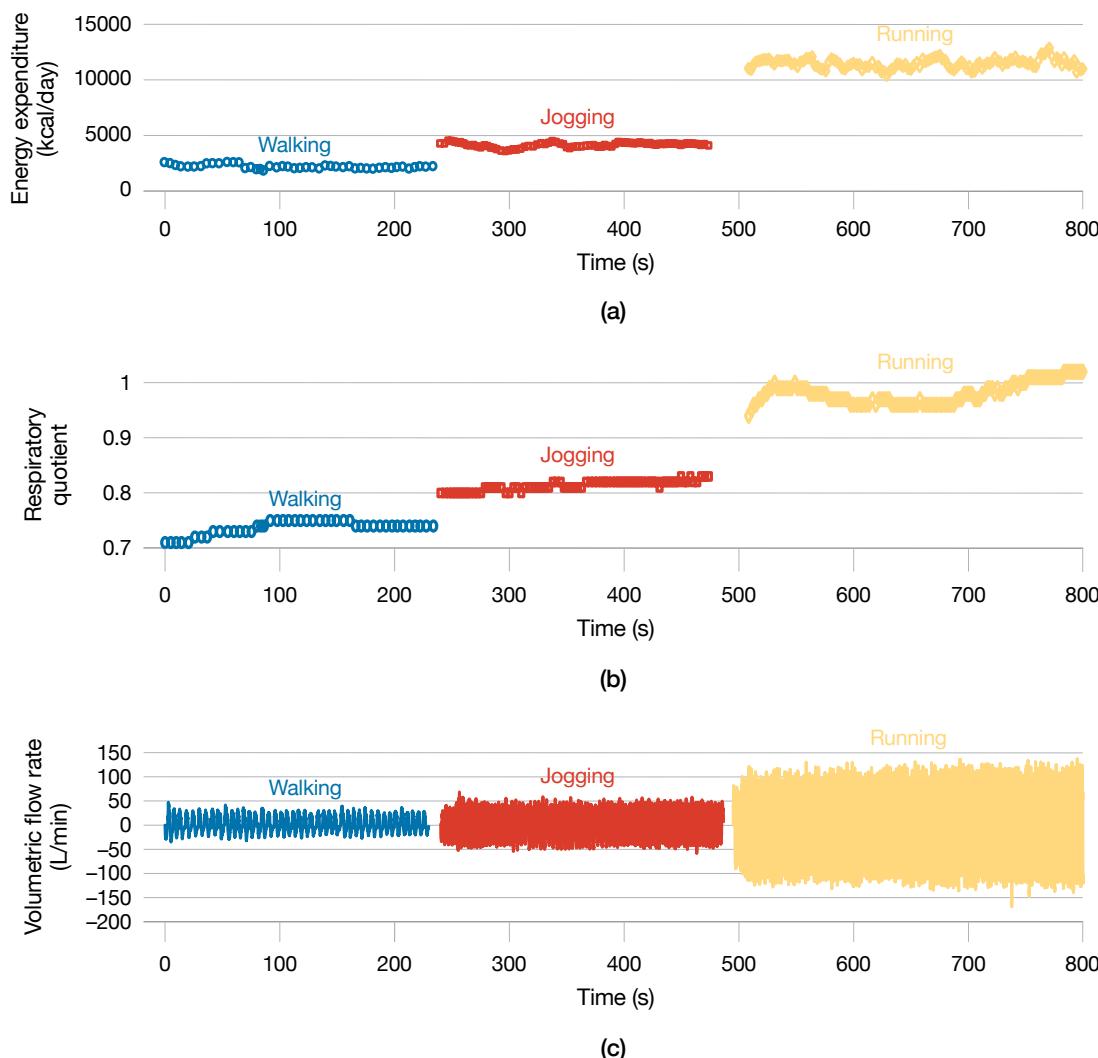
FIGURE 13. Three different sample sets of data were collected with the COBRA sensor. Walking resulted in a respiratory quotient (RQ) ~0.75 and energy expenditure (EE) ~2200 kcal/day (a); jogging resulted in RQ ~0.80 and EE ~4400 kcal/day (b); and running resulted in RQ ~1 and EE ~12,000 kcal/day (c). The changes in RQ reveal how metabolic fuel preference shifts to glucose in response to the higher-intensity exercise.

rate in the third. The flow measurement is performed at 18–20 hertz to provide the resolution needed to resolve the breath profiles, and the other measurements are presented in the form of average numbers over the course of a whole breath. The data presented as breath-by-breath data are not truly breath-by-breath measurements because the time constants associated with the mixing chamber and gas sensors are larger than the duration of an average breath. Consequently, the RQ and energy expenditure data represent a moving average over two to eight adjacent breaths, depending upon respiration rate and tidal volume.

One important aspect to observe across the three exercise intensities is the change in the dynamic range of the flow and energy estimates. For walking flow measurements, the peak flow rate is around 30 liters/minute; however, as the exercise intensity increases, the peak flow rates increase to upwards of 150 liters/minute. In general, the peak flow rate may increase to more than 400 liters/minute during maximal exertion and can be less than 10 liters/minute for resting measurements of individuals with small lung capacity. Designing a reliable passive proportional sampling scheme for this large dynamic range is a challenge. To ensure accuracy and rapid speed of measurements at both ends, the current prototype employs a low-volume-rate flow tube for resting and low-intensity exercise measurements, and a high-volume-rate flow tube for moderate to intense exercise or exertion.

The metabolic response to each of these three different levels of exercise intensity is evident in these data. In the case of walking, the RQ is dominated by slow twitch muscular processes, which are typically fat burning. In this particular case, the subject has an RQ of 0.75, or a molar mix of 17 percent carbohydrates and 83 percent fats. As the subject transitions to jogging, more glycogen from the skeletal tissue is used for energy production, and the RQ increases to 0.80 while the EE increases from about 2200 to 4400 kcal. As the exercise intensity further increases to a run, glucose becomes the preferred fuel and dominates the macronutrient selection, raising the RQ to 1. These data exemplify the transition in metabolic fuel preference as a function of exercise intensity. The transition points and the values of RQ and EE, paired with the exercise information, can be used to assess individual fitness and track changes in fitness and endurance over time.

The U.S. Army Research Institute of Environmental Medicine has conducted a rigorous independent evaluation of the COBRA sensor with 12 volunteers. The testing included a series connection between the COBRA sensor and a gold-standard metabolic cart that ensured that the same test subject's metabolism was presented to both sensors during a measurement cycle. Results of this testing indicated that over a range of activity levels from resting to walking, jogging, and running, the COBRA produces EE and RQ measurements consistent with those of the reference metabolic cart [66].

Applications
TRAINING AND PERFORMANCE TRACKING

For high-performance athletes or military personnel, optimally matching dietary macronutrients to the demands of a specific activity can improve performance and increase the likelihood of successful completion of the activity. In endurance athletic events, matching nutrition to performance needs is already standard practice in the form of carbohydrate or glycogen loading [58], but this practice often results in overconsumption of carbohydrates to ensure maximization of glycogen stores, with excess carbohydrates being converted to fat and stored.

However, to better predict an individual's available energy resources for longer time periods and to properly account for the activity demands and dietary intake, a metabolic model is necessary. Dietary tracking and exercise simulation can be employed to estimate an athlete's initial energy reserve state. The COBRA sensor can then be used to monitor which nutrients are actually selected to meet metabolic energy needs and estimate the individual's future nutritional needs. The information available from the metabolic fuel model and simulation may be used alone or in conjunction with other physiological models to improve performance predictions and assess the benefits of training. One possible example of coupling metabolic and physiological models is estimating a distance runner's performance. A 2010 study used the level of exertion and glycogen storage state to predict the maximal running distance of a person [67]. The COBRA sensor reports both VO_2 and RQ, which are the correlated measures of exertion recommended in the study [67], and the metabolic fuel model predicts glycogen storage on the basis of previous diet and exercise. The information

provided by the sensor and the model allows individuals training for an endurance event to predict how far they can expect to run at different running speeds and intensities. The COBRA sensor data are invaluable for guiding different training regimens or deciding on an optimal pace to run a race.

This reasoning can be extended to address needs in the military. For example, a mission may require demanding physical exertion over the course of many days in a location where resupply is challenging. In such a situation, it would be beneficial for decision makers to know which soldiers operate efficiently on calorie-dense fatty food and which soldiers require high-carbohydrate food to maintain the required performance levels. Measurements of RQ and EE during training sessions provide the information for commanders to know which soldiers preferentially select which type of fuel substrate when executing various tasks and levels of exertion. With this knowledge, the commander can select soldiers exhibiting a preference for energy from fat oxidation for missions with difficult resupply. Conversely, for short-lived intense physical tasks, a soldier may need to be capable of exerting maximal metabolic output, in which case a diet that assures adequate glycogen stores is essential. The best way to optimize soldier selection and nutrition for the job at hand is to measure each soldier's metabolic profile and capacities.

The U.S. Army's recruitment and retention has historically relied on a one-size-fits-all fitness assessment of candidates that is independent of their eventual roles. Soldier 2020 is an Army program taking a more discerning approach, with the goal of developing specific fitness standards based on the role that the individual fills [68]. These standards will be informed with awareness of the metabolic costs of particular jobs and quantitative metabolic measurements of the soldiers who successfully occupy the roles. The Soldier 2020 system will likely ease fitness requirements for some specialties while still maintaining the rigor and high standards currently in place for the necessary roles. For instance, a recruit who wants to work in the field still requires a high level of physical fitness, whereas soldiers working at desk jobs could have their standards reduced. The job-specific fitness requirements could help commanders better retain and recruit high-functioning soldiers for specific needs.

QUANTIFYING INDIVIDUAL METABOLISM

In addition to improving military training and readiness, the COBRA sensor and metabolic fuel model can be employed to improve the metabolic health of the civilian population on a broad scale. Specific to the military, the physical health and well-being of soldiers have come under increased scrutiny as the rate of obesity in the military has monotonically increased over the past several years [69]. The rise of obesity has raised the washout rate of soldiers unable to meet fitness requirements and has also elevated health care expenses [16]. Studies have shown that health maintenance, rather than intervention after a person becomes pathogenic, mitigates the risk of such events. However, implementing a health maintenance regimen can be challenging, especially to ensure that individuals follow prescribed diets and exercise regimens. Additionally, very little feedback is available to allow individuals to know if they are doing enough, or even doing the right thing, to achieve weight loss or weight maintenance goals. A system model to predict the impact of dietary and exercise choices on the body and a sensor to provide on-demand feedback and confirm progress are previously unavailable tools for planning and confirming health maintenance and weight management regimens.

A typical practice for weight management studies or intervention is to record activity levels, food consumption, and body weight over the course of weeks or months. Personal tracking and estimation of food calories is challenging to do accurately, and body weight is highly variable, requiring days or weeks to confirm the impact of day-to-day diet and exercise choices. Body weight measurements can be deceptive, as body mass composition depends on water retention, muscle mass, fat, glycogen stores, digestive state, and many other difficult-to-quantify variables. The metabolic fuel model helps to visualize the consequences of eating a certain macronutrient mix or engaging in a particular exercise rather than trying to quantify feeling better or tracking confounding weight changes. Because the metabolic fuel model incorporates macronutrient type and quantity, dietary frameworks (e.g., low fat versus low carbohydrate) can be utilized to gain an understanding of the impact of specific macronutrients and exercise regimens rather than simply tracking calories apart from macronutrient composition. While weight change is often tracked as the

only measure of success, our sensor provides on-demand feedback of metabolic fuel substrate to indicate whether the subject is staying in fat-burning mode or fat-storing mode. Furthermore, the model shows how blood insulin and glucose levels (quantities that are linked to feelings of hunger or satiation) are impacted by macronutrient and exercise choices [70, 71]. The COBRA sensor reliably evaluates metabolic indicators and provides a more complete context to help validate the model and understand how dietary choices are affecting metabolism on a day-to-day or even hour-by-hour basis [61].

If, rather than dietary advice, a specific metabolic state—such as ketosis (extreme restriction of carbohydrates) or glycogen loading (extreme consumption of carbohydrates)—is sought, a measurement of the RQ informs the users where they are on the carbohydrate versus fat fuel substrate spectrum. If their goal is to move toward a ketogenic state, an RQ measurement will establish whether fat is the preferred metabolic fuel or if the body is relying primarily on carbohydrates to meet energy demands. On the other end of the spectrum, a resting measurement of RQ greater than 1 would indicate that fat burning is being suppressed in favor of carbohydrate metabolism and that excess carbohydrates are being converted into fats.

Future Work
FIELD TESTING

As noted earlier, the traditional methods of metabolic measurement, whole room indirect calorimetry or treadmill testing with a metabolic cart, are expensive and do not reflect the actual conditions under which soldiers operate in the field. In conjunction with the metabolic fuel model, providing COBRA sensors to soldiers in the field will enable the collection of metabolic data over a range of workloads and environmental conditions. The COBRA sensor data will help quantify the impact of environmental factors on energy expenditure and better inform the development of doctrine and training manuals to avoid excessive heat strain, glycogen depletion, and other undesirable performance degradations.

SINGLE-SUBJECT EXPERIMENTS AND CLINICAL STUDIES

The low-size, -weight, and -power metabolic sensor, along with other physiological status monitors such as heart rate and continuous glucose monitors, promises to enhance the value of clinical studies by providing a low-cost, simple-to-use sensor for monitoring RQ and EE in natural living conditions. For example, a handheld metabolic sensor, communicating with a camera-equipped smartphone to document meal composition and an activity monitor to record exercise, enables the collection of a much richer dataset for clinicians and researchers to use in quantifying the impact of diet and exercise on metabolic health.

There are a number of single-subject experiments planned to provide a preliminary assessment of the value of the metabolic sensor in helping individuals achieve various metabolic fitness and health goals.

EARLY TYPE 2 DIABETES DETECTION AND INTERVENTION

Prediabetes is defined by a fasting blood glucose level between 100 and 125 mg/dL. According to the metabolic model described earlier in this article, a normally functioning endocrine system strives to keep blood sugar below 100 to 110 mg/dL at all times. A primary mechanism for achieving this regulation in the face of carbohydrate overconsumption is to defer the oxidation of fats and preferentially oxidize glucose to meet metabolic energy needs. Consequently, a persistently high resting RQ, particularly in the post prandial, or fasting state, would seem to be indicative of excessively high blood glucose levels. If so, routine resting RQ measurement may provide an early indication of prediabetes. In the case of confirmed prediabetic or type 2 diabetic individuals, rather than making multiple capillary blood glucose measurements each day to check for high blood sugar, a noninvasive RQ measurement in the middle range may be sufficient to ensure blood glucose levels are not high and thus avoid the need to draw a capillary blood sample for testing.

The efficacy of RQ measurements as a noninvasive means of identifying high blood glucose levels can be tested by instrumenting subjects with a continuous glucose monitor, intentionally raising blood sugar levels through manipulation of dietary macronutrients, and quantifying the correlation between high resting RQ and high blood glucose levels. A study with 16 subjects examined this thesis using a whole room indirect calorimeter to collect data for 24 hours [61]. Demonstration of a consistent correlation over a longer period of time would justify a larger study to establish

the consistency of the correlation between resting RQ and blood glucose levels over a range of ages, gender, activity, and dietary habits.

GUIDANCE FOR WEIGHT MANAGEMENT PROTOCOLS

The traditional approach to weight management is to estimate calories associated with food intake and activity, create an energy-deficient diet, and monitor progress toward weight loss goals with a scale. Not surprisingly, this approach fails much of the time for several reasons. As noted earlier, counting food calories and estimating activity calories is fraught with error, and body weight may fluctuate by several pounds or more a day depending upon hydration state, glycogen stores, and digestive state.

Rather than a one-size-fits-all formulaic CICO diet for weight loss, the metabolic model and sensor provide the quantitative tools to tailor macronutrient intake and exercise activity to increase the likelihood of achieving weight loss and weight management goals. In particular, a quantitative measurement of RQ may prevent dieters from deceiving themselves about the success of efforts to balance dietary macronutrients and the effectiveness of their exercise regimen by providing real-time feedback. The COBRA sensor enables on-demand RQ measurement throughout the day to assess the impact of dietary and exercise choices on the goal of staying in the fat-burning zone.

RQ-ONLY SENSOR

At least half of the volume of the COBRA sensor is devoted to making the flow measurements necessary to compute energy expenditure and related flow metrics, such as tidal volume and VO_2. However, considering the spectrum of potential applications for the metabolic fuel model, knowledge of energy expenditure is not always necessary. In particular, for detecting chronically high blood glucose levels and providing guidance on how dietary macronutrient imbalances impact fuel substrate selection, energy expenditure data are not required.

By removing the requirement to measure volumetric flow rates and focusing only on resting RQ, a number of simplifications can be made to the COBRA sensor, the breath sampling protocol, and the processing. The flow tube can be made smaller, and the four tubes connecting the flow tube to the mixing chamber can be reduced to a single tube. Because the RQ is a ratio of the measured CO_2

concentration to O_2 concentration for a fixed volume, it is no longer necessary to capture all of the exhaled breath, eliminating the need for a nose clip. The best measure of RQ comes from the deep alveolar lung exchange, so it is only necessary to sample the end tidal portion of the breath. Consequently, an RQ-only sensor can be made smaller and in a form factor compatible with connection to a smartphone for battery power, processing, and display of the measured data.

An RQ-only sensor would reduce the cost of manufacturing and the operational power, further reducing the barriers to personal ownership by a broad segment of the population.

Acknowledgments

References

1. H.A. Krebs, "The Citric Acid Cycle," Nobel Lecture, 1953.
2. M.L. Vetter, S.J. Sharon, J. Herring, M. Sood, N.R. Shah, and A.L. Kalet, "What Do Resident Physicians Know about Nutrition? An Evaluation of Attitudes, Self-Perceived Proficiency, and Knowledge," *Journal of the American College of Nutrition*, vol. 27, no. 2, 2008, pp. 287–298.
3. Centers for Disease Control and Prevention, "National Health and Nutrition Examination Survey," 18 June 2018, available at https://www.cdc.gov/nchs/nhanes/index.html.
4. K.M. Adams, K.C. Lindell, M. Kohlmeier, and S.H. Zeisel, "Status of Nutrition in Medical Schools," *American Journal of Clinical Nutrition*, vol. 83, no. 4, 2006, pp. 941S–944S.
5. R.D. Feinman, W.K. Pogozelski, A. Astrup, R.K. Bernstein, E.J. Fine, E.C. Westman, et al., "Dietary Carbohydrate Restriction as the First Approach in Diabetes Management: Critical Review and Evidence Base," *Nutrition*, vol. 31, no. 1, 2015, pp. 1–13.
6. O. Hamry, "Nutrition Revolution—The End of the High Carbohydrates Era for Diabetes Prevention and Management," *U.S. Endocrinology*, vol. 10, no. 2, 2014, pp. 103–104.
7. A.L. Mayen, P. Marques-Virdal, F. Paccaud, P. Bovet, and S. Stringhini, "Socioeconomic Determinants of Dietary Patterns in Low- and Middle-Income Countries: A Systematic Review," *The American Journal of Clinical Nutrition*, vol. 100, no. 6, 2014, pp. 1520–1531.

8. U. Schwab, L. Lauritzen, T. Tholstrup, T.I. Haldorsson, U. Riserus, M. Uusitupa, and W. Becker, "Effect of the Amount and Type of Dietary Fat on Cardiometabolic Risk Factors and Risk of Developing Type 2 Diabetes, Cardiovascular Diseases, and Cancer: A Systematic Review," *Food and Nutrition Research*, vol. 58, no. 1, 2014.

9. C.B. Ebbeling and D.S. Ludwig, "The Carbohydrate-Insulin Model of Obesity," *Journal of the American Medical Association Internal Medicine*, vol. 178, no. 8, 2018, pp. 1098–1103.

10. K.D. Hall, S.J. Guyenet, and R.L. Leibel, "The Carbohydrate-Insulin Model of Obesity is Difficult to Reconcile with Current Evidence," *Journal of the American Medical Association Internal Medicine*, vol. 178, no. 8, 2018, pp. 1103–1105.

11. Centers for Disease Control and Prevention, "National Center for Health Statistics Obesity Data," 29 Mar. 2017, available at https://www.cdc.gov/chronicdisease/resources/publications/aag/pdf/2016/diabetes-aag.pdf.

12. W. Youfa and M.A. Beydoun, "The Obesity Epidemic in the United States—Gender, Age, Socioeconomic, Racial/Ethnic, and Geographic Characteristics: A Systematic Review and Meta-regression Analysis," *Epidemiologic Reviews*, vol. 29, no. 1, 2007, pp. 6–28.

13. B. Christeson, K. Clifford, A.D. Taggart, and B. Goodman, "Too Fat, Frail, and Out-of-Breath to Fight," Mission: Readiness, 23 Apr. 2015, available at http://mission-readiness.s3.amazonaws.com/wp-content/uploads/MN-Physical-Activity-Report.pdf.

14. J. Levy, L.M. Segal, K. Thomas, R. St. Laurent, A. Lang, and J. Rayburn, "F as in Fat: How Obesity Threatens America's Future," Trust for America's Health/Robert Wood Johnson Foundation, 16 Aug. 2013, available at https://www.rwjf.org/en/library/research/2013/08/f-as-in-fat--how-obesity-threatens-america-s-future-2013.html.

15. K.S. Mudambo, C.M. Scrimgeour, and M.J. Rennie, "Adequacy of Food Rations in Soldiers During Exercise in Hot, Day-Time Conditions Assessed by Doubly Labelled Water and Energy Balance Methods," *European Journal of Applied Physiology and Occupational Physiology*, vol. 76, no. 4, 1997, pp. 346–351.

16. Y. Wei, A. Ravelo, T.H. Wagner, C.S. Phibbs, A. Bhandari, S. Chen, and P.G. Barnett, "Prevalence and Costs of Chonic Conditions in the VA Health Care System," *Medical Care Research and Review*, vol. 60, supp. 3, 2003, pp. 146S–167S.

17. D. Keener, K. Goodman, A. Lowry, S. Zaro, and L.K. Khan, "Recommended Community Strategies and Measurements to Prevent Obesity in the United States: Implementation and Measurement Guide," *Centers for Disease Control and Prevention Morbidity and Mortality Weekly Report*, vol. 58, no. 7, 2009, pp. 1–26.

18. J.H. Ledikwe, J.A. Ello-Martin, and B.J. Rolls, "Portion Sizes and the Obesity Epidemic," *The Journal of Nutrition*, vol. 135, no. 4, 2005, pp. 905–909.

19. L.R. Young and M. Nestle, "The Contribution of Expanding Portion Sizes to the U.S. Obesity Epidemic," *American Journal of Public Health*, vol. 92, no. 2, 2002, pp. 246–249.

20. S.O. McDoniel, "A Systematic Review on Use of a Handheld Indirect Calorimeter to Assess Energy Needs in Adults and Children," *International Journal of Sport Nutrition and Exercise Metabolism*, vol. 17, no. 5, 2007, pp. 491–500.

21. R.E. Keesey and M.D. Hirvonen, "Body Weight Set-Points: Determination and Adjustment," *The Journal of Nutrition*, vol. 127, no. 9, 1997, pp. 1875S–1883S.

22. R.B. Harris, "Role of Set-Point Theory in Regulation of Body Weight," *The Federation of American Societies for Experimental Biology Journal*, vol. 4, no. 15, 1990, pp. 1875S–1883S.

23. T.C. Ni and M.A. Savageau, "Model Assessment and Refinement Using Strategies from Biochemical Systems Theory: Application to Metabolism in Human Red Blood Cells," *Journal of Theoretical Biology*, vol. 179, no. 4, 1996, pp. 329–368.

24. H. Kitano, "Computational Systems Biology," *Nature*, vol. 420, no. 7912, 2002, pp. 206–210.

25. P. James and R. McFadden, "Understanding the Processes Behind the Regulation of Blood Glucose," *Nursing Times*, vol. 100, no. 16, 2004, pp. 56–58.

26. C. Ruggiero and L. Ferrucci, "The Endeavor of High Maintenance Homeostasis: Resting Metabolic Rate and the Legacy of Longevity," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 61, no. 5, 2006, pp. 466–471.

27. S.T. Chung, S.K. Chacko, A.L. Sunehag, and M.W. Haymond, "Measurements of Gluconeogenesis and Glycogenolysis: A Methodological Review," *Diabetes*, vol. 64, no. 12, 2015, pp. 3996–4010.

28. A.J. Pantophlet, S. Wopereis, C. Elderink, R.J. Vonk, J.H. Stroeve, S. Biijlsma, et al., "Metabolic Profiling Reveals Differences in Plasma Concentrations of Arabinose and Xylose after Consumption of Fiber-Rich Pasta and Wheat Bread with Differential Rates of Systemic Appearance of Exogenous Glucose in Healthy Men," *The Journal of Nutrition*, vol. 147, no. 2, 2016, pp. 152–160.

29. D. Zeevi, T. Korem, D. Israeli, D. Rothschild, A. Weinberger, O. Ben-Yacov, et al., "Personalized Nutrition by Prediction of Glycemic Responses," *Cell*, vol. 163, no. 5, 2015, pp. 1079–1094.

30. A.E. Jeukendrup and G.A. Wallis, "Measurement of Substrate Oxidation During Exercise by Means of Gas Exchange Measurements," *International Journal of Sports Medicine*, vol. 26, supp. 1, 2005, pp. S28–S37.

31. F. Peronnet and D. Massicotte, "Table of Non-protein Respiratory Quotient: An Update," *Canadian Journal of Sport Sciences*, vol. 16, no. 1, 1991, pp. 23–29.

32. S.P. Brown, W.C. Miller, and J.M. Eason, *Exercise Physiology: Basis of Human Movement in Health and Disease*. Philadelphia: Lippincott Williams & Wilkins, 2006.

33. T.P. Stein, R.W. Hoyt, M. O'Toole, J.J. Leskiw, M.D. Schluter, R.R. Wolfe, and W.D.B. Hiller, "Protein and Energy Metabolism During Prolonged Exercise in Trained Athletes," *International Journal of Sports Medicine*, vol. 10, no. 5, 1989, pp. 311–316.

34. Boundless and Lumen Learning, "Boundless Biology," 2018, available at https://courses.lumenlearning.com/boundless-biology/.

35. G.A. Binns and R. Di Brezzo, "Thermic Effect of Food, Exercise, and Total Energy Expenditure in Active Females," *Journal of Science and Medicine in Sport*, vol. 18, no. 2, 2015, pp. 204–208.

36. E.F. Coyle, A.R. Coggan, M.K. Hemmert, and J.L. Ivy, "Muscle Glycogen Utilization During Prolonged Strenuous Exercise When Fed Carbohydrate," *Journal of Applied Physiology*, vol. 61, no. 1, 1986, pp. 165–172.

37. D.R. Seals, J.M. Hagberg, W.K. Allen, B.F. Hurley, G.P. Dalsky, A.A. Ehsani, and J.O. Holloszy, "Glucose Tolerance in Young and Older Athletes and Sedentary Men," *Journal of Applied Physiology*, vol. 56, no. 6, 1984, pp. 1521–1525.

38. K.J. Acheson, Y. Schutz, T. Bessard, K.R. Anantharman, J.P. Flatt, and E. Jequier, "Glycogen Storage Capacity and de Novo Lipogenesis During Massive Carbohydrate Overfeeding in Man," *The American Journal of Clinical Nutrition*, vol. 48, no. 2, 1988, pp. 240–247.

39. J.A. Romijn, E.F. Coyle, L.S. Sidossis, A. Gastaldelli, J.F. Horowitz, E. Endert, and R.R. Wolfe, "Regulation of Endogenous Fat and Carbohydrate Metabolism in Relation to Exercise Intensity and Duration," *American Journal of Physiology—Endocrinology and Metabolism*, vol. 265, no. 3, 1993, pp. E380–E391.

40. P. Knuiman, M.T. Hopman, and M. Mensink, "Glycogen Availability and Skeletal Muscle Adaptations with Endurance and Resistance Exercise," *Nutrition and Metabolism*, vol. 12, no. 59, 2015.

41. M.M. McGrane, *Carbohydrate Metabolism: Synthesis and Oxidation*. Philadelphia: WB Saunders, 2000.

42. L. Sokoloff, "Metabolism of Ketone Bodies by the Brain," *Annual Review of Medicine*, vol. 24, 1973, pp. 271–280.

43. F. Ameer, L. Scandiuzzi, S. Hasnain, H. Kalbacher, and N. Zaidi, "De Novo Lipogenesis in Health and Disease," *Metabolism*, vol. 63, no. 7, 2014, pp. 895–902.

44. M.K. Hellerstein, "De Novo Lipogenesis in Humans: Metabolic and Regulatory Aspects," *European Journal of Clinical Nutrition*, vol. 53, supp. 1, 1999, pp. S53–S65.

45. D. Mathias, "How Obesity Causes Type 2 Diabetes," in *Staying Healthy From 1 to 100: Diet and Exercise, Current Medical Knowledge on How to Keep Healthy*. Berlin: Springer, 2011, p. 46.

46. M. Lazar, "How Obesity Causes Diabetes: Not a Tall Tale," *Science*, vol. 307, no. 21, 2005, pp. 373–375.

47. J. Ye, "Emerging Role of Adipose Tissue Hypoxia in Obesity and Insulin Resistance," *International Journal of Obesity*, vol. 33, no. 1, 2009, pp. 54–66.

48. E.C. Westman, W.S. Yancy, J.C. Mavropoulos, M. Marquart, and J.R. McDuffie, "The Effect of a Low-Carbohydrate, Ketogenic Diet versus a Low-Glycemic Index Diet on Glycemic Control in Type 2 Diabetes Mellitus," *Nutrition and Metabolism*, vol. 5, no. 36, 2008.

49. L.R. Saslow, A.E. Mason, S. Kim, V. Goldman, R. Ploutz-Snyder, H. Bayandorian, et al., "An Online Intervention Comparing a Very Low-Carbohydrate Ketogenic Diet and Lifestyle Recommendations versus a Plate Method Diet in Overweight Individuals with Type 2 Diabetes: A Randomized Controlled Trial," *Journal of Medical Internet Research*, vol. 19, no. 36, 2017, p. E36.

50. J.V. Nielsen and E.A. Joensson, "Low-Carbohydrate Diet in Type 2 Diabetes: Stable Improvement of Bodyweight and Glycemic Control During 44 Months Follow-Up," *Nutrition and Metabolism*, vol. 5, no. 14, 2008.

51. P. Dyson, "Low Carbohydrate Diets and Type 2 Diabetes: What Is the Latest Evidence?" *Diabetes Therapy*, vol. 6, no. 4, 2015, pp. 411–424.

52. U. Ozcan, Q. Cao, E. Yilmaz, A. Lee, N.N. Iwakoshi, E. Ozdelen, et al., "Endoplasmic Reticulum Stress Links Obesity, Insulin Action, and Type 2 Diabetes," *Science*, vol. 306, no. 5695, 2004, pp. 457–461.

53. K.N. Frayn, "Calculation of Substrate Oxidation Rates in Vivo from Gaseous Exchange," *Journal of Applied Physiology*, vol. 55, no. 2, 1983, pp. 628–634.

54. J.D. Weir, "New Methods for Calculating Metabolic Rate with Special Reference to Protein Metabolism," *The Journal of Physiology*, vol. 109, no. 1–2, 1949, pp. 1–9.

55. J.H. Goedecke, A.S. Gibson, L. Grobler, M. Collins, T.D. Noakes, and E.V. Lambert, "Determinants of the Variability in Respiratory Exchange Ratio at Rest and During Exercise in Trained Athletes," *American Journal of Physiology—Endocrinology and Metabolism*, vol. 279, no. 6, 2000, pp. E1325–E1334.

56. K. Wasserman, B.J. Whipp, S.N. Koyl, and W.L. Beaver, "Anaerobic Threshold and Respiratory Gas Exchange During Exercise," *Journal of Applied Physiology*, vol. 35, no. 2, 1973, pp. 236–243.

57. M. Sun, G.W. Reed, and J.O. Hill, "Modification of a Whole Room Indirect Calorimeter for Measurement of Rapid Changes in Energy Expenditure," *Journal of Applied Physiology*, vol. 76, no. 6, 1994, pp. 2686–2691.

58. Parvo Medics, "Parvo Medics TrueOne 2400," available at http://www.parvo.com/trueone-2400/.

59. COSMED—The Metabolic Company, cosmed.com, available at http://www.cosmedusa.com/en/.

60. L. Jonge, T. Nguyen, S.R. Smith, J.J. Zachwieja, H.J. Roy, and H.J. Bray, "Prediction of Energy Expenditure in a Whole Body Indirect Calorimeter at Both Low and High Levels of Physical Activity," *International Journal of Obesity*, vol. 25, no. 7, 2001, pp. 929–934.

61. A. Gribok, J.L. Leger, M. Stevens, R. Hoyt, M. Buller, and W. Rumpler, "Measuring the Short-Term Substrate Utilization Response to High-Carbohydrate and High-Fat Meals in the Whole-Body Indirect Calorimeter," *Physiological Reports— American Physiological Society*, vol. 4, no. 12, 2016.

62. L.M. Candell, C. Ferraiolo, G.A. Shaw, A.M. Siegel, and G. Zogbi, "Systems, Apparatus, and Methods Related to Modeling, Monitoring, and/or Managing Metabolism," U.S. patent application no. 2017/0055875 A1, issued 2 Mar. 2017.

63. L.M. Candell, C. Ferraiolo, G.A. Shaw, A.M. Siegel, G. Zogbi, H.L. McClung, and R.W. Hoyt, "Systems, Apparatus, and Methods Related to Modeling, Monitoring, and/or Managing Metabolism," U.S. patent no. 10,638,956, issued 5 May 2020.

64. H. McClung, R.W. Hoyt, L.M. Candell, J.M. Mahan, G.A. Shaw, A.M. Siegel, R.L. Standley, and K.J. Thompson, "Passive, Proportional Measurement of Oxygen and Carbon Dioxide Consumption for Assessment of Metabolic Parameters," U.S. patent application no. 2020/0022618 A1, issued 23 Jan. 2020.

65. K. Thompson, G.A. Shaw, A.M. Siegel, and L.M. Candell, "Methods and Apparatus for Passive, Proportional, Valveless Gas Sampling and Delivery," U.S. patent application no. 2020/0022617 A1, issued 23 Jan. 2020.

66. H.L. McClung, L.A. Walker, D.P. Looney, W.J. Tharion, A.P. Welles, H.M. Hansen, et al., "Validity of a New Portable Metabolic Gas Exchange System," *Official Journal of the American College of Sports Medicine*, vol. 51, no. 5, 2019, p. S745.

67. B.I. Rapoport, "Metabolic Factors Limiting Performance in Marathon Runners," *Public Library of Science Computational Biology*, vol. 6, no. 10, 2010.

68. M.C. Koester, "Physical Demands Study Goal: Identify the Ideal Soldier for the Job," *Non-Commissioned Officers Journal*, 2015.

69. L.L. Clark and S.B. Taubman, "Update: Diagnoses of Overweight and Obesity, Active Component, U.S. Armed Forces, 2011–2015," *Medical Surveillance Monthly Report*, vol. 23, no. 9, 2016, pp. 9–13.

70. M.I. Grossman and I.F. Stein, "Vagotomy and the Hunger-Producing Action of Insulin in Man," *Journal of Applied Physiology*, vol. 1, no. 4, 1948, pp. 263–269.

71. J. Rodin, "Insulin Levels, Hunger, and Food Intake: An Example of Feedback Loops in Body Weight Regulation," *Health Psychology*, vol. 4, no. 1, 1985, pp. 1–24.

About the Authors

Lawrence M. Candell is an assistant head of the Space Systems and Technology Division at Lincoln Laboratory. He specializes in signal processing, electro-optical systems, and optical communications. He has contributed to cutting-edge imaging systems ranging from gigapixel cameras for persistent surveillance to million-frame-per-second cameras for analyzing missile defense missions. He has developed not only ground-based advanced sensor prototypes but also sensor prototypes for air, rocket, and space platforms. As the associate program manager for the Mars Laser Communications Demonstration, he was responsible for the development of a distributed aperture receiving system that could successfully decode with efficiencies of nearly three bits/photon. He has been a key leader across the entire spectrum of the Space Systems and Technology Division's system and technology programs. He holds bachelor's and master's degrees in electrical engineering and a bachelor's degree in management, all from MIT.

Gary A. Shaw is a senior staff member in the Integrated Systems and Concepts Group. He develops low-power wireless sensors and architectures, including photon-counting communication systems and metabolic modeling and sensing systems, to aid soldiers. He served 12 years as an assistant and associate group leader at the Laboratory and three years as the test director and assistant leader of the ALCOR C-band and Millimeter Wave radars, which are located at the Reagan Test Site in the Marshall Islands. As a senior staff member, he has worked on novel architectures and sensors to enable high-performance sensing and communication with low size, weight, power, and cost systems. He holds bachelor's and master's degrees in electrical engineering from the University of South Florida and a doctoral degree in electrical engineering from the Georgia Institute of Technology, where he taught for one quarter before accepting a position at Lincoln Laboratory.

Kyle J. Thompson is a member of the technical staff in the Mechanical Engineering Group. He joined Lincoln Laboratory in 2015, and his research focuses on developing metabolic modeling and sensing systems and building low-temperature apparatus and devices. He holds bachelor's degrees in mathematics and physics from the University of Massachusetts and a doctoral and master's degree in physics from the University of Florida.

Noncontact Laser Ultrasound for Biomedical Imaging Applications

Robert W. Haupt, Charles M. Wynn, Matthew R. Johnson, Jonathan R. Fincke, Xiang Zhang,

Brian W. Anthony, and Anthony E. Samir

Lincoln Laboratory, the MIT Medical Electronic Device Realization Center, and Massachusetts General Hospital are developing an optical system that acquires ultrasound images of the human body's interior without touching the patient. The noncontact laser ultrasound (N-CLUS) images tissue and bone and measures their elastographic properties from a distance of several inches to meters from the patient. The N-CLUS can be automated to provide a repeatable fixed frame of reference and could compete with magnetic resonance and computed tomography while reducing the system cost, complexity, and health risks. Such a system would expand the role of accurate imaging capabilities inside and outside the hospital center.

>> **Ultrasound is an ideal method used in** medical practice to image the human body's interior [1]. It is portable, inexpensive, and causes no harmful biological effects in the patient while producing images with excellent resolution. Moreover, the ultrasound capability can measure blood flow and provides information about the mechanical properties (elastography) of tissue, organs, and bone. Despite ultrasound's advantages, magnetic resonance imaging (MRI) and computed tomography (CT) are the dominant imaging methods in medicine even though they impose health risks and are greater than ultrasound systems in size, cost, and complexity. Ultrasound would be ubiquitous in all of medicine but for one central problem: the need for transducer-to-skin contact necessitates manual device operation, which introduces variability between measurements, reduces the availability of ultrasound systems to patients, and increases the cost of the system. Manual transducer operation has been a central limitation of medical ultrasound for 50 years.

Contact transducers suffer problems in irreproducibility of measurements because each individual operator spontaneously chooses the transducer's look angle and application pressure to couple and transmit ultrasonic waves into the patient. These spontaneous actions differ greatly between measurements and between different medical operators, resulting in interoperator variability. This variability causes unpredictable signal and image distortion. More specifically, the localized tissue's mechanical properties change because of compaction differences in the vicinity of the transducer. Compaction occurs when the operator pushes the transducers onto

the patient's skin to achieve signal coupling, causing the skin to bunch together or distort. The unevenness of the skin causes variable ultrasonic wave velocities and travel paths. Theses numerous variations distort the clarity of the measured features' geometry and dimensions at depth as displayed in the resulting ultrasound image. Because of the variability between measurements, repeat measurements can miss important changes in tissue features over time. Ultrasound is further weakened by the lack of a fixed frame of reference to register the image location. Thus, MRI and CT are typically mandated over ultrasound because these methods provide fixed, gantry-mounted transmit-and-receive systems that enable spatially reproducible images and permit more accurate comparison of disease states over time.

Interoperator variability greatly limits ultrasound applications. For example, the American Thyroid Association's ultrasound guidelines require an increase in thyroid nodule volume of more than 50 percent—or an increase of more than 20 percent in two or more measured dimensions—before a biopsy can be recommended [2]. Large increases in nodule size are necessary before nodule growth can be confidently diagnosed because of the high interoperator variability associated with conventional ultrasound. The European Organization for Research and Treatment of Cancer and Response Evaluation Criteria in Solid Tumors' guidelines [3] do not permit the use of ultrasound tumor measurements for the follow-up of cancer because of high interoperator variability. Patients with cancer must be followed up with CT or MRI scans.

The advantages of three different medical imaging methods are compared in Table 1. The CT method offers high-resolution imaging capabilities that have been used for many years. The MRI method sets the standard for volumetric imaging inside the human body to detect and diagnose a variety of maladies in tissue, organs, the brain, and bone without the radiation risk of CT. Moreover, MRI techniques have evolved for new applications, such as tumor detection (elastography), in which anomalous voxels (3D pixels) in an MRI image indicate stiffer tissue masses that could be a sign of cancer. However, because MRI and CT systems are large, expensive, and require highly trained medical staff to operate, they are impractical for mass production and for many field-forward settings (locations outside hospital centers). A

fundamental drawback to MRI and CT is that they are only available to a small number of patients per day because measurement with these devices take a long time (30 to 60 minutes per patient) and because there are few units per hospital. There is great appeal in developing accurate imaging systems that combine the spatial reproducibility advantages of CT and MRI with the low-cost, portability, ease of operation, high-throughput, and low-toxicity advantages of medical ultrasound.

Appeal of Noncontact Laser Ultrasound

Noncontact laser ultrasound (N-CLUS) is designed to mitigate operator variability while providing a fixed-reference measurement capability [4]. Unlike contact transducers, laser ultrasound does not cause large-scale tissue compaction, and it also eliminates the variable of transducer look angle selection. Laser excitation and receive locations on the skin's surface can be revisited with spatial precision, permitting the system to monitor subtle changes in the geometry of measured features and the mechanical properties of tissue over time. Another advantage of laser ultrasound is that it does not necessitate contact with damaged or burned skin, painful regions, traumatized tissue, areas susceptible to infection and contamination, and areas awkward or difficult to reach. Standoff laser ultrasound can be directed with a mirror and is less likely to interfere with other medical procedures, such as surgery, compared to conventional ultrasound. With the proper development, N-CLUS medical imaging systems could potentially compete with MRI and CT as a lower-cost, portable, and health risk–free alternative. Such systems would expand the role of accurate imaging capabilities beyond the hospital center and enable numerous new medical ultrasound applications. We are currently receiving funding from the U.S. Army to develop N-CLUS for critical life-saving systems that are used to assess soldiers' injuries in field-forward environments. The MRI and CT methods are highly impractical for such environments because of their large weight, size, and power requirements.

The N-CLUS system generates an ultrasonic wave by first emitting short, high-peak-power infrared pulses from a Q-switched laser. The short pulses of optical energy reach the skin's surface and convert to mechanical energy via thermal expansion in the shallow layers of the

ROBERT W. HAUPT, CHARLES M. WYNN, MATTHEW R. JOHNSON, JONATHAN R. FINCKE, XIANG ZHANG, BRIAN W. ANTHONY, AND ANTHONY E. SAMIR

Table 1. The Pros and Cons of Magnetic Resonance Imaging and Computerized Tomography Are Compared to the Pros and Cons of Current Medical Ultrasound

TYPE OF IMAGING METHOD	PROS	CONS
MAGNETIC RESONANCE IMAGING (MRI) 	• High resolution • Safe • Fixed-reference gantry—enables time image progression monitoring	• Large, expensive, in-hospital • Danger from metal debris • Time-consuming • Uncomfortable for many patients—can require sedation in some cases
COMPUTERIZED TOMOGRAPHY (CAT SCAN) 	• High resolution • Fixed-reference gantry—enables time image progression monitoring	• Large, expensive, in-hospital • Weak on tissue contrasts • Radiation risk • Time-consuming • Uncomfortable for many patients—can require sedation in some cases
ULTRASOUND 	• High resolution • Portable, inexpensive • Safe • In-hospital and field-forward	• Image distortion from interobserver variability • No fixed reference • Impractical for time progression imaging in many applications
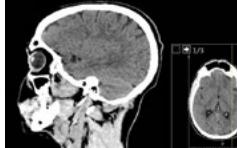
skin; then, the optically absorptive material of the skin rapidly deforms and launches a propagating mechanical wave response [5]. These processes are termed photoacoustic effects. From our studies, we found that long infrared wavelengths absorb rapidly in tissue (the wave attenuates in tissue depths of less than 1 millimeter) and create a consistent uniform acoustic source that generates a coherent broadband ultrasonic wave that propagates well into the far field within tissue (several inches there and back). We observed that this acoustic source is not influenced by skin pigment or topography, which

indicates that such a system could be used on any patient or skin condition. Finally, the mechanical/vibrational signature of the ultrasonic wave returning from inside the body is measured at the skin's surface with a laser Doppler vibrometer (LDV).

Research on noncontact acoustic and vibrational ultrasound has been ongoing for the past two decades [6–11]. Much of the focus has been on photoacoustic tomography (PAT), which is a popular method used in studies to image near-surface shallow capillaries in animal tissue. In PAT, pulsed infrared light penetrates

tissue and induces photoacoustic phenomena preferentially in chromophores (e.g., hemoglobin). Contact transducers then measure the emitted acoustic response. The PAT approach employs near-infrared wavelengths (700 to 900 nanometers) to maximize the penetration of light to the depth of interest and map optical absorptivity variations in tissue, such as blood vessels, with excellent spatial resolution. However, optical scatter severely limits the penetration of PAT to tissue depths of less than 1 centimeter. Recent studies have explored replacing contact transducers with an LDV as a sensing device to make the PAT system totally noncontact [10, 12]. However, for optical measurement systems to compete with practiced medical ultrasound, these systems need to generate acoustic waves in the frequency bands relevant to clinical ultrasound while being able to propagate to depths of several inches in the body. The laser system would need to operate within skin- and eye-safe laser powers and provide submillimeter resolution, an excellent signal-to-noise ratio (SNR), and rapid data acquisition times.

Operational Concept: Standoff N-CLUS Wave Generation with Laser Measurement

In the N-CLUS system, photoacoustic excitation is used in conjunction with an LDV to acquire ultrasonic waves without touching the patient's skin, as demonstrated in Figure 1. Photoacoustic phenomena develop from optical energy shaped in very short (picoseconds to nanoseconds) laser pulses that interact with an optically absorptive target surface, such as water or biological tissue, in which optical energy rapidly converts into heat [5, 15]. This nearly instantaneous heating creates a concentration of mechanical stress within the irradiated tissue patch. Because the tissue is no longer in mechanical equilibrium, the stress dissipates and generates a propagating acoustic wave into the tissue mass at a high frequency of approximately 1 megahertz.

Photoacoustic Phenomena Leading to Ultrasonic Wave Propagation in Tissue

Photoacoustic phenomena involve the conversion of optical energy into a localized pressure rise in biological tissue and are well described in the literature [6, 13–15]. The Gruneisen parameter, Γ, of tissue relates the initial pressure, p_0, to the light absorption in the expression

$$p_0 = \Gamma \mu_a F, \text{ where } \Gamma = \beta v_s^2 / C_p$$

where μ_a is the optical absorption coefficient of tissue, F is the local light fluence (optical intensity), β is the isobaric volume expansion coefficient, v_s^2 is the acoustic wave speed in tissue, and C_p is the specific heat of the tissue. The local pressure rise then results in acoustic wave propagation and is expressed below as a function of distance and time (r, t).

$$\nabla^2 p(\mathbf{r},t) - \frac{1}{v_s^2}\frac{\partial^2 p(\mathbf{r},t)}{\partial t^2} = \frac{\beta}{C_p}\left[\mu_a + \Delta\mu_a(\mathbf{r})\right]\frac{\partial F(\mathbf{r},t)}{\partial t}$$

Photoacoustic phenomena and the process of ultrasonic wave propagation in biological tissue are explicitly derived and shown in detail in this article's Appendix.

Laser Doppler Vibrometry Sensitivity and Noise-Floor Components

To construct a fully noncontact laser ultrasound system, we used a heterodyne coherent laser radar (ladar) that acted as a vibrometer to measure the returning vibrational signals of ultrasonic waves emerging at the tissue's surface. In heterodyne detection, a signal of interest at a certain frequency is nonlinearly mixed with a reference local oscillator set at a nearby frequency. The desired output is the difference signal, which carries the information (amplitude, phase, and frequency modulation) of the original higher frequency signal. Optical photons are detected by their energy levels or by photon counting at the receiver and are proportional to the square of the electric field, thus forming a nonlinear event. When the local oscillator and the signal beams impinge together on a target surface, they mix, producing heterodyne beat frequencies.

The sensitivity performance of the laser Doppler vibrometer (LDV) is determined by its noise floor [16]. The noise floor is the level of unwanted signals, or noise, that a signal must surpass to be detected. The total noise floor is expressed below.

$$N_{total} = \sqrt{\left(N_{shot}\right)^2 + \left(N_{speckle}\right)^2 + \left(P, N_{patient}\right)^2}$$

where N is noise and P is platform.

ROBERT W. HAUPT, CHARLES M. WYNN, MATTHEW R. JOHNSON, JONATHAN R. FINCKE, XIANG ZHANG, BRIAN W. ANTHONY, AND ANTHONY E. SAMIR

FIGURE 1. The diagrams demonstrate the notional N-CLUS system and how it acquires ultrasound images and measures mechanical properties of the body's interior without touching the subject. An optical head containing an excitation source and laser-sensing receiver is scanned over the body (a). The optical source emits pulsed infrared light, which enters the skin's surface layers and creates ultrasonic waves. The ultrasonic returns reflected off of the body's internal structures return to the surface of the skin and are measured with a laser Doppler vibrometer (b).

The shot noise is receiver-power limited, speckle noise is induced by motion, and platform and patient noise entail vibrations other than the ultrasonic waves of interest.

Shot noise depends on the number of photoelectrons received per second, ϕ_e (optical return from the vibrating tissue surface), over the vibrometer's demodulated bandwidth, and the shot noise in turn determines the received signal quality. The shot-noise spectrum of the surface particle velocity, $A_{v,sh}$, as a function of frequency, f, is proportional to the received returning photoelectrons for the transmit optical wavelength, λ, such that

$$A_{v,sh}(f) = \frac{f\lambda}{\sqrt{\phi e}}$$

As the bandwidth and desired frequency increase, the need for more photoelectrons also increases to maintain the desired noise floor. Increasing the LDV's fluence level provides more photoelectrons but runs the risk of exceeding skin- and eye-safety limits. The possibility of exceeding these limits is a critical concern in the development of a laser ultrasound system for patient skin exposure.

Speckle is the noise that occurs because of the distribution of optical scatterers on the tissue's surface that a laser beam encounters. For a diffuse surface, there are likely many optical scatterers (because of surface roughness) that reflect light back to the LDV. The speckle noise contribution to the laser-sensing system can be reduced by signal time integration with respect to the same realization of scatterers. The frequency-dependent speckle noise amplitude is shown below.

$$A_{v,sp}(f) = \lambda \sqrt{\frac{\pi f_{exc}^2}{12}} \sqrt{\frac{2\alpha}{\alpha^2 + (2\pi f)^2}}$$

where $\alpha = 2\pi f_{exc}$, and $f_{exc} = v_t/d$ (beam translation velocity on target over the beam diameter) is the exchange rate of the speckle pattern.

In terms of the LDV performance for ultrasonic measurements, the shot-noise contribution (at 1 megahertz) is anticipated to dominate the noise-floor sensitivity. When any motion from the patient is introduced to the system, speckle noise becomes a significant factor. Even subtle involuntary patient motion can produce significant fluctuations in the speckle realization and resultant noise floor. The design of a successful noncontact laser receiver critically depends on shot- and speckle-noise sources being sufficiently reduced to achieve a useful return SNR of the ultrasonic wave.

Lincoln Laboratory N-CLUS Demonstration System

We developed and built a proof-of-concept N-CLUS system to explore the phenomenology controlling optical absorption that generates useful ultrasonic waves and to demonstrate how to measure the waves without tissue contact. We explored the variety of elastic/ultrasonic waves that can be generated from a photoacoustic source, constructed images of tissue interiors with various systems, and compared these images to those produced by standard medical ultrasound. We explored the N-CLUS data-acquisition configurations that measure different ultrasonic wave types, such as longitudinal and shear waves in bone, which we analyzed to determine the bone's elastographic properties (elastic moduli) in vivo.

To provide the N-CLUS photoacoustic signal, we used several commercially available Q-switched laser sources, including a variable Continuum Panther optical parametric oscillator that drives the photoacoustic excitation source at prescribed optical wavelengths from the deep ultraviolet to near infrared, and a customized 1550-nanometer eye- and skin-safe power source. A fast steering mirror was used to position the laser beam onto the tissue sample's surface within submillimeter increments in the vertical and horizontal directions. A Polytec RSV-150 LDV was used as the receive system; its beam could be positioned with or without a mirror to the receiving location. The Polytec employs an eye- and skin-safe 1550-nanometer wavelength transmission at 10 milliwatts.

Return ultrasonic time series were collected on a Tektronix 200-megahertz digitizing scope and then recorded on a laptop computer. The excitation beam's location on the tissue sample was controlled by the fast steering mirror, which was positioned using an automated computer routine written in LabView software. Typically, scan lines are steered across the sample. A single scan line forms a 2D cross-sectional profile of the tissue sample. Multiple scan lines can be compiled to form a time image that is processed, converted to show depth, and migrated to display in 3D the physical layout of the tissue's interior features.

We employed a bistatic (i.e., set at different locations, as opposed to collocated) source and receiver geometry while varying the separation distance to accommodate standard seismic data processing algorithms that remove the effects of time distortion, such as normal move out and diffraction of point scatterers [18]. Normal move out corrections and time migration [17] are used to reduce the plotted cross-sectional data to obtain depth images of the tissue sample's interior. The N-CLUS experimental apparatus and measurement configuration are shown in Figure 2.

ROBERT W. HAUPT, CHARLES M. WYNN, MATTHEW R. JOHNSON, JONATHAN R. FINCKE, XIANG ZHANG, BRIAN W. ANTHONY, AND ANTHONY E. SAMIR
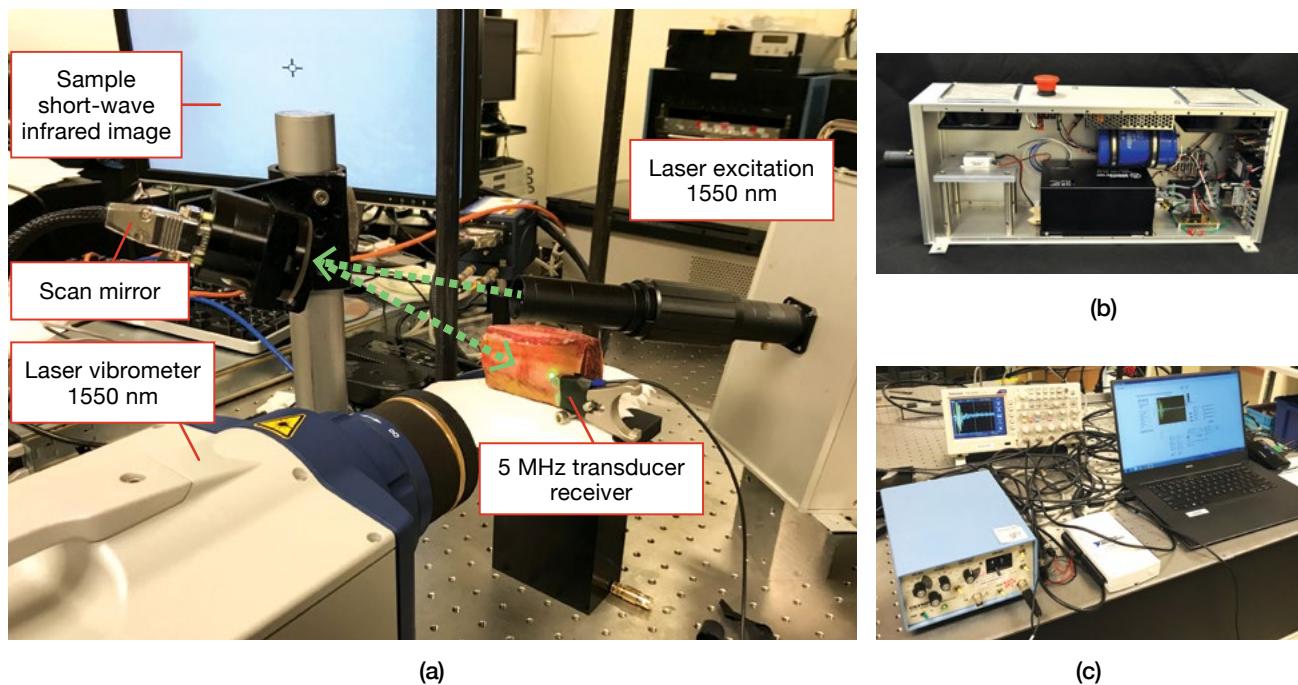
FIGURE 2. The N-CLUS proof-of-concept system (a) uses a customized photoacoustic source emitting a 1550-nanometer eye- and skin-safe pulse for excitation and a commercial laser Doppler vibrometer (LDV) to measure the returning ultrasonic wave. The LDV receive system is a Polytec RSV-150. A fast steering mirror scans and positions the optical beams on the biological tissue sample. The custom-built Lincoln Laboratory 1550-nanometer Q-switched pulsed laser (b) induces the photoacoustic effect to send an ultrasonic wave into tissue. The laser has a pulse repetition frequency (PRF) of 10 hertz and 1 millijoule of power per pulse. The N-CLUS approach's data acquisition system (c) comprises a Tektronix scope and a laptop. The mirror assembly can be controlled with LabView software to scan and position the excitation laser beam onto a tissue sample.

N-CLUS Eye- and Skin-Safe Operation and Performance

Critical factors to consider when one is designing an operational system for use by physicians are patient safety and risks related to system hazards; image quality; effective aerial coverage rates; size, weight, and power; and cost. We first evaluated the effects of an imaging system's optical excitation wavelength on the SNR, image quality, and skin and eye safety. To start, we examined the image quality produced by four different common optical excitation wavelengths that span the near- to short-wave infrared—810 nanometers, 1064 nanometers, 1550 nanometers, and 2000 nanometers—using the optical parametric oscillator at Lincoln Laboratory. For each of these wavelengths, a pulse with a 7-nanosecond rise time (the time it takes for a pulse to rise from its lowest value to its highest value) was transmitted; the fluence level of the laser beam on the sample's surface was held constant at 21 millijoules/square centimeter

(mJ/cm^2), and the laser spot on the sample's surface was 2 millimeters in diameter. To obtain the measurements, we used a beef steak sample that contained two 1-millimeter-diameter rods that were inserted lengthwise into the sample. The pulsed laser beam that launches the ultrasonic wave was scanned across the sample, and the LDV measured the returning waves from a location at the far end of the scan line. The data collected from this demonstration were then displayed to show a 2D cross section of a time series where the y-axis showed the two-way travel time of the ultrasonic wave and the x-axis showed the scan position of the excitation laser where the ultrasonic wave was launched.

In Figure 3b, the measured ultrasonic time-domain reflection data are compared in sonograms. As the optical wavelength increased, the observed ultrasonic frequency band also shifted higher relative to the optical absorption. Correspondingly, the sonograms showed ultrasonic reflection events sharpening with increasing

optical wavelength excitation. The ultrasonic amplitudes increased with increasing optical wavelengths up to the 1550-nanometer source, then showed a relative decrease for the 2000-nanometer source. Although the ultrasonic frequency content generated by the 2000-nanometer source is high and the associated image exhibits the best temporal resolution, the signal amplitude drops. This drop is likely caused by increased ultrasonic wave attenuation acting more severely on higher-frequency components as the ultrasonic wave traveled through the beef sample.
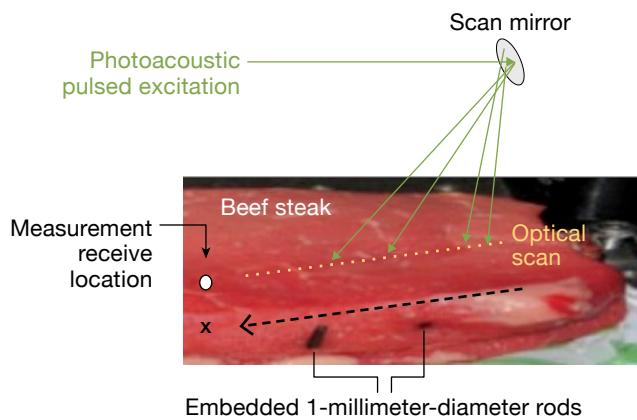
(a)
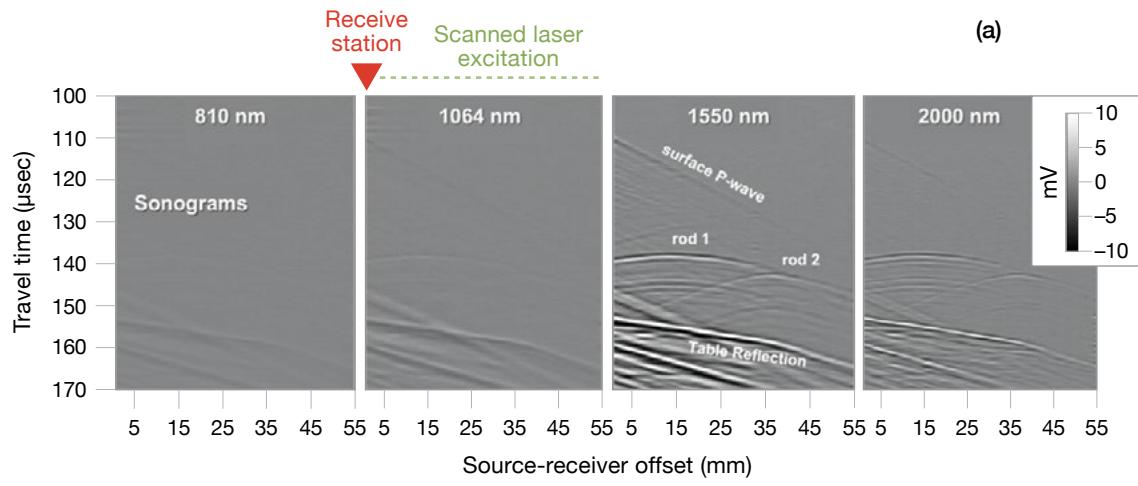
(b)
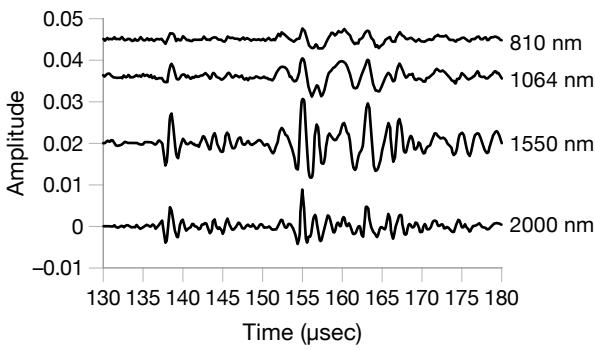
(c)
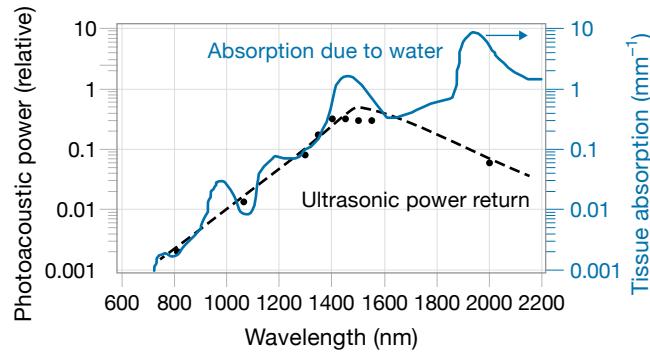
(d)

FIGURE 3. A beef steak sample is shown with the noncontact laser ultrasound geometry layout (a). Two 1-millimeter-diameter rods were embedded in the sample, and the pulsed laser was scanned across the sample by using a steering mirror. The scan line in this geometry is referred to as an off-end sonogram, where the receiver station is at the near end of the cross section, and the laser ultrasound excitation source is moved incrementally away from the receiver in a bistatic geometry. Noncontact laser ultrasound sonograms were produced for four stimulus wavelengths with a constant spot size and power per area (b). The 1550-nanometer source displayed the best image resolution. Also shown is a comparison of ultrasonic time-series traces for measurements directly over a 1-millimeter rod at four optical wavelengths (c). The 1550-nanometer and 2000-nanometer sources registered the rod the most clearly, as indicated by the amplitude level. On the bottom right, the chart shows the summed ultrasonic power of measured ultrasound returns as a function of optical excitation wavelength, and the optical absorption caused by water (the dominant material in biological tissue) as a function of optical wavelength (d).

ROBERT W. HAUPT, CHARLES M. WYNN, MATTHEW R. JOHNSON, JONATHAN R. FINCKE, XIANG ZHANG, BRIAN W. ANTHONY, AND ANTHONY E. SAMIR

Table 2 states the American National Standards Institute (ANSI)'s laser exposure limits for eye and skin safety for the four wavelength examples shown in Figure 3b [18]. The 1550-nanometer and 2000-nanometer sources produced images with the best SNR. The images showed the features of the metal rods, steak, and table top while maintaining optical excitation within skin- and eye-safety limits. The 810-nanometer and 1064-nanometer sources produced weaker SNR images for optical excitation fluence levels above the ANSI eye-safety threshold. The high optical absorptivity of the 1550-nanometer wavelength provides the highest eye-safety threshold, allowing high optical powers that improve the SNR while still maintaining safety. The 1550-nanometer source also offers common commercially ready components. Of the four wavelengths examined, the 1550-nanometer and 2000-nanometer sources provide the highest ultrasound image quality and SNR while providing the highest margin for safety.

N-CLUS Imaging in Biological Tissue

We next examine N-CLUS's ability to generate ultrasonic images using skin-and eye-safe laser powers in biological tissue and compare the system's performance to that of contact transducers and a GE 9-megahertz medical ultrasound probe. First, we scanned grocery store beef steak and pork samples with the N-CLUS 1550-nanometer photoacoustic source and used the Polytec RSV-150 vibrometer to measure the returning ultrasonic waves at the surface of each sample. We then used the N-CLUS system to generate anatomical images of tissue and bone in the forearm of a human test subject. To simulate more realistic patient conditions, particular attention was given to these measurements to ensure that the emitted optical powers were within skin- and eye-safety requirements.

Comparison of Medical Transducer Signal Quality to N-CLUS Signal Quality

If the N-CLUS approach is to lead to a viable medical system, it must produce high-quality interior images of tissue that are comparable to images acquired with conventional medical ultrasound. We first compare the signal with noise of N-CLUS to that of medical transducer technologies, which are shown in Figures 4 and 5. For the demonstration shown in Figure 4, we constructed

Table 2. ANSI Z136.1-2007 Skin and Eye Safety Optical Exposure Levels

WAVELENGTH (nm)	SKIN SAFETY THRESHOLD (mJ/cm^2)	EYE SAFETY THRESHOLD (mJ/cm^2)
810	33	8.3×10^{-4}
1064	105	5.0×10^{-3}
1550	1000	1000
2000	100	100

a simulated tissue sample with a PVC pipe embedded in a copolymer soft tissue material that had the same mechanical properties as human tissue. Compared methodologies included an all-acoustic transducer transmit and transducer receive system in a water tank [19]; a hybrid-acoustic optical photoacoustic transmit with acoustic contact transducer receive system; and an all-optical, totally noncontact laser ultrasound transmit-and-receive system.

In the water tank, the simulated tissue's surface was at a distance of 4 centimeters from the all-acoustic transmit-and-receive transducer configuration. The transducers were immersed in water to mitigate the effects of varying transducer application pressures and look angles. The three configurations depicted in Figure 4 show the signal-with-noise characteristics of each method for acquiring ultrasound data. All of the techniques show similar SNR for their time-series trace measurements of the returning ultrasonic signal. The traces shown for the all-acoustic (immersed in water tank) and the hybrid-acoustic (optical photoacoustic source using the contact transducer attached to the simulated tissue) methods exhibited the same approximate SNR for a single transmitted pulse response. This result indicates that the optical photoacoustic source is significantly efficient and is able to achieve an SNR that is reasonably comparable to that of the conventional all-acoustic transducer system. We conclude from this example study that the photoacoustic source used at 1550 nanometers is potentially competitive, in terms of efficiency and quality, with a single medical ultrasound transducer.

In the third plot in Figure 4, the N-CLUS-acquired trace is shown in comparison to the results of the other

two methods. The laser vibrometer's measurement was averaged over 128 repeated pulse response collections to improve the SNR. The vibrometer performance is driven by its shot noise and speckle noise. To be competitive with conventional all-acoustic transducer systems, the laser-sensing capability of N-CLUS requires significant averaging.

Note that only the all-acoustic transducer approach used the water tank, which required transducers to be submerged. Optical methods are immersion free, require a line of sight in the air, and can be steered by a mirror.

Although all-acoustic water tank ultrasound can provide high-quality data and images, the water tank itself has serious drawbacks for field-forward applications because of its size, weight, and complexity. Furthermore, the water tank is often uncomfortable for patients to immerse their body parts in for measurement. It is highly impractical to immerse the entire torso, so measurements are limited to the extremities.

The measurements demonstrated in Figure 4 show that a photoacoustically generated ultrasonic wave in tissue can be comparable in magnitude and SNR to
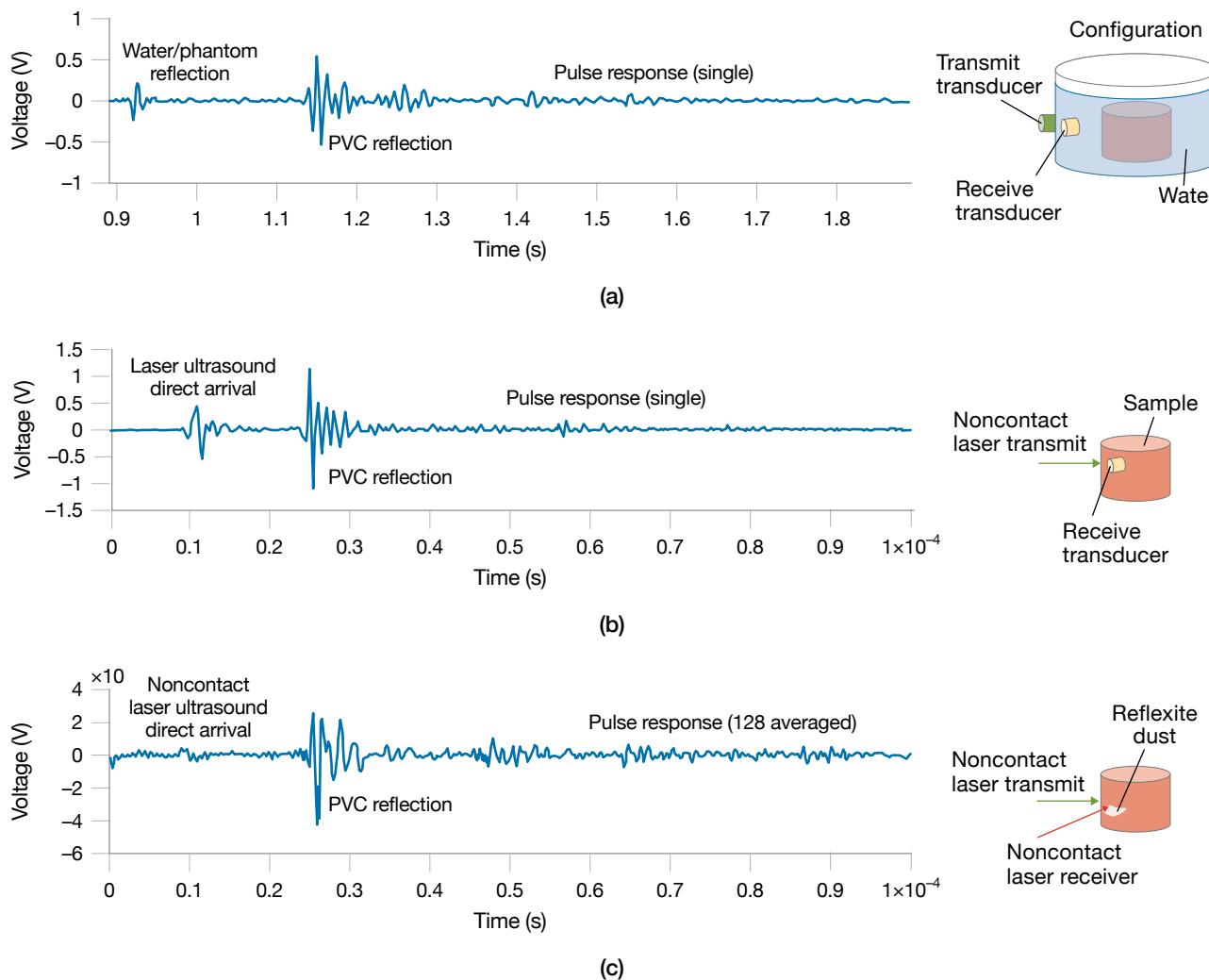
FIGURE 4. We compared time-series traces measured by state-of-the-art all-acoustic piezoelectric transducers submerged in a water tank (a), a hybrid-acoustic optical laser photoacoustic excitation measured by a contact Olympus V152 longitudinal wave 1-megahertz transducer (b), and an all-optical total noncontact laser ultrasound (c). The noncontact ultrasound's time-series trace closely resembles those of the ultrasound transducers, indicating that waves generated by the noncontact system are comparable in magnitude and SNR to waves generated by traditional transducers.

ROBERT W. HAUPT, CHARLES M. WYNN, MATTHEW R. JOHNSON, JONATHAN R. FINCKE, XIANG ZHANG, BRIAN W. ANTHONY, AND ANTHONY E. SAMIR

waves produced by medical ultrasound transducers. We next constructed 2D sonograms of a complex soft tissue mass containing bone to compare the skin- and eye-safe N-CLUS system to optically generated ultrasound (photoacoustic) that utilizes contact transducers (1-megahertz Olympus V152 longitudinal wave). In the demonstration shown in Figure 5, a 3-inch-thick beef steak sample containing a rib bone was scanned with the full N-CLUS system (10-hertz PRF, 1550-nanometer optical-to-ultrasound excitation, and 1550-nanometer Polytec laser vibrometer receiver) to produce an ultrasound image. The full N-CLUS image was compared with the image produced by a hybrid acquisition method that paired photoacoustic excitation with a contact ultrasound transducer. Each time-series trace used in each image had a listen time of 200 microseconds, which provided enough time to capture the ultrasonic signals-of-interest returns from the steak sample. The laser vibrometer and transducer were filtered to a bandwidth of 1.5 megahertz for comparison. The beef sample had no optical treatment to enhance the laser vibrometer's performance. A water-based medical gel was used to enhance coupling of the transducer to the beef sample's surface (using gel is a standard medical practice in ultrasound). The emitted optical beams were within skin- and eye-safety limits.

The geometry of the beef sample's internal features are well captured in both images. The fat layers and bone are clearly evident because the traveling ultrasonic wave encounters more resistance at these features, thus causing a strong acoustic impedance signature. We next compared ultrasound images acquired with the N-CLUS technique to images acquired with a state-of-the-art General Electric (GE) 9-megahertz medical ultrasound system, as shown in Figure 6. In this example, the photoacoustic source and LDV beams were collocated and scanned across the sample.

In both the N-CLUS and GE Array images, the fat layer geometries are clearly evident. However, the spatial resolution is significantly better for the ground truth GE system's image. The lower resolution of the N-CLUS image is primarily because the set bandwidth of the receiving laser vibrometer is limited at 1 megahertz. The GE system also employs array-processing methods to further improve the spatial resolution. It is anticipated that increasing the laser vibrometer bandwidth
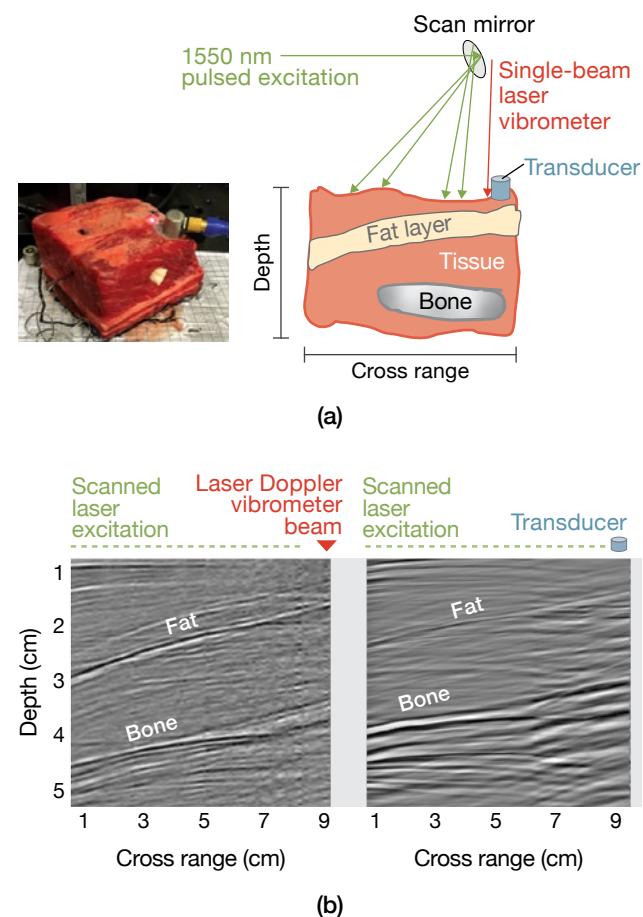
(a)

(b)

FIGURE 5. A 3-inch-thick beef steak sample containing a rib bone was scanned with the full N-CLUS system (photoacoustic source with stationary LDV receiver) and then with a hybrid system that paired a photoacoustic source with an Olympus V152 longitudinal wave 1-megahertz transducer (a). The transducer and LDV were approximately collocated. The sonograms produced by both systems are shown converted to depth (b). Both the total N-CLUS system (left) and the scanned photoacoustic source with the transducer (right) clearly displayed the fat and bone features within the steak sample.

by a factor of nine would improve the N-CLUS system's image resolution and bring it closer to the resolution of the GE system.

At present, the N-CLUS system is much slower than the commercial GE system in terms of data acquisition. The largest factor impeding the N-CLUS data acquisition rate is the very slow 10-hertz PRF of the 1550-nanometer optical excitation laser that generates the ultrasonic pulse. In Figure 6, 100 excitation positions constituted the data acquisition scan, and the scan required 10 seconds for each pulse excitation. The total time was increased even
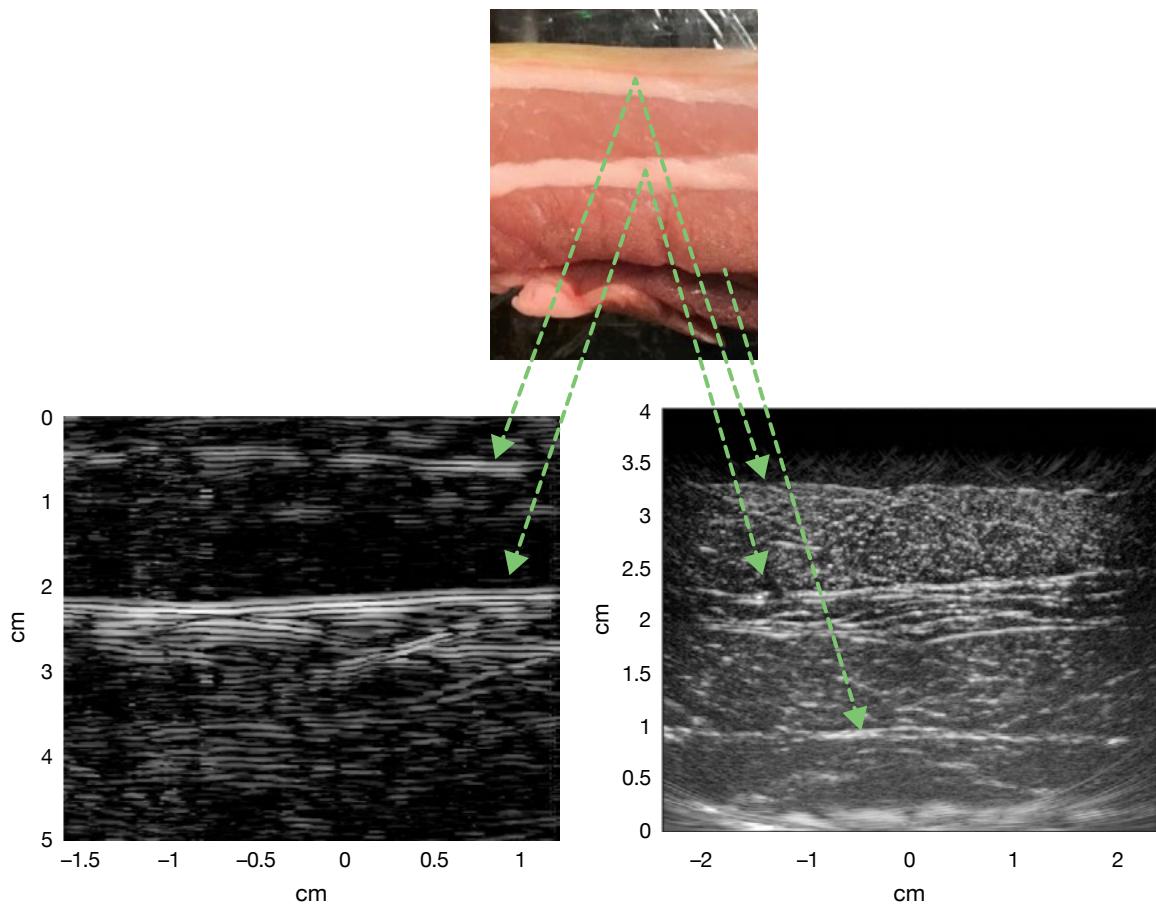
FIGURE 6. The above figure shows the N-CLUS sonograms of a scanned pork sample, converted to depth. Comparisons are shown for the total N-CLUS system (collocated scanned photoacoustic source and scanned laser Doppler vibrometry beam) (left) and the GE 9-megahertz medical ultrasound system (right). The fat layers of the pork sample can be seen in the sonograms for both systems.

further because acquiring the N-CLUS image involved averaging 64 data-trace collections per scan location, and it took approximately 10 minutes to acquire the entire dataset. The laser vibrometer required averaging to yield a useful SNR for the returning signal.

N-CLUS Eye- and Skin-Safe Images of a Human Test Subject

By using the N-CLUS and GE 9-megahertz systems, we acquired 2D cross-sectional ultrasound images of a human test subject's forearm. For these image constructions, the N-CLUS optical excitation and laser vibrometer beams were slightly offset (about 0.5 centimeters apart) and then scanned together across the test subject's skin. The N-CLUS optical powers were consistently maintained to be skin and eye safe. No optical treatment was applied to the surface of the test subject's skin to enhance the

laser vibrometer's performance. However, for the GE 9-megahertz system demonstration, a water-based gel was applied to the subject's skin to enhance ultrasound coupling during operation of the GE 9-megahertz transducer head.

In both the N-CLUS and GE Array images, fat layers, a tendon, and bone features are clearly evident. However, the spatial resolution is again significantly better for the GE system's image. The N-CLUS image has a lower resolution because, as previously mentioned, the set bandwidth of the receiving laser vibrometer is limited at 1 megahertz. The N-CLUS system required less than five minutes to acquire the entire dataset in the examples shown in Figure 7.

Another important factor to consider when using the N-CLUS system to collect data is patient motion. It becomes increasingly difficult for a patient to remain

ROBERT W. HAUPT, CHARLES M. WYNN, MATTHEW R. JOHNSON, JONATHAN R. FINCKE, XIANG ZHANG, BRIAN W. ANTHONY, AND ANTHONY E. SAMIR

(a)

(b)

(c)

(d)

FIGURE 7. This graph shows the results of a time-series data measurement acquired with a laser Doppler vibrometer (a). The spikes in amplitude show the direct ultrasonic surface wave and the subsequent reflections off tissue and bone in the test subject's arm. The subject's arm was scanned in the radial and axial directions (b). The green spot is an eye- and skin-safe reference laser that marked the position of the infrared 1550-nanometer laser beam, which is not visible to the naked eye. The N-CLUS sonograms were converted to depth [27] for scans of the subject's forearm in the axial direction (c) and the radial direction (d). Comparisons are shown for the total N-CLUS system (collocated scanned photoacoustic source and scanned LDV beam) (left) and the GE 9-megahertz medical ultrasound system (right). Internal features of the arm, such as tissue layers, a tendon, and bone, can be seen in the sonograms for both systems.

motionless for periods longer than several seconds. Patient motion, such as sudden jerks or readjustment to become comfortable, results in partial shifts in the resulting image or static breaks causing misalignment between trace collections when they are averaged. To overcome the effects of patient motion on image quality, the N-CLUS system's data-acquisition time needs to be significantly shortened. Many patient-motion effects can be mitigated by using higher-PRF optical excitation systems. Optical short-pulse systems (5-nanosecond pulses operated at a

PRF of 5 kilohertz) that are available on the commercial market can also provide optical powers on the order of one to several millijoules per pulse. At the appropriate optical wavelength, these powers are skin and eye safe. The fast PRF can decrease the total time needed for data acquisition by a factor of 500. Applying a fast excitation PRF to the N-CLUS system reduced its data-acquisition time from 1 to 10 minutes to approximately 1 second. Because the time required to take a measurement is much shorter when a short PRF is applied, the effects of patient motion will be greatly reduced, enabling the N-CLUS scanning approach to be more feasible than before for operational and practical ultrasound.

The results discussed in this section are a first in the medical literature. We produced anatomical images by using skin- and eye-safe laser ultrasound within test tissue samples without using surface treatments to enhance the optical return for a laser vibrometer measurement. Moreover, this is a first in the medical literature in performing such measurements successfully on a human test subject and producing a useful anatomical image [20–24].

Use of N-CLUS for Elastography Measurements and Novel Applications

Elastography is an emerging field in medicine that noninvasively measures the mechanical properties and determines the spatial distribution of tissue, organs, and bones in a patient's body. Evaluation of elastographic properties entails quantifying the stiffness distribution of tissue. The degree of stiffness and compliance (the ability of an organ to increase its volume with pressure) indicates tissue type and health. For example, stiffening of liver tissue can imply fibrosis of the liver, which is a serious health condition. Elastography is increasingly being used for investigating disease conditions in internal organs. It can also be used as diagnostic information to supplement anatomical images. Elastography can be used to guide biopsies or to replace them entirely. Biopsies are invasive and painful and present a risk of hemorrhage or infection, whereas elastography is completely noninvasive. Elastography is now commonly used for the detection and diagnosis of breast, thyroid, and prostate cancers. Certain types of elastography are also suitable for musculoskeletal diagnosis and can determine the mechanical properties and states of muscles and tendons.

To produce anatomical images, we exploited far-field ultrasonic waves that reflect and transmit within soft tissue in the body's interior. We present a novel approach to acquiring the elastographic properties of bone with the N-CLUS system, which can reveal the elastic moduli by using far-offset surface waves that travel within the outer surface layers of bone. Far-offset surface waves in bone can be induced and measured by using a bistatic configuration in which the photoacoustic source and laser vibrometer receiver are widely separated. The diving wave travels along the tissue-bone interface and re-emerges at a long offset, as depicted in Figure 8a.

The surface longitudinal and shear wave events are exhibited simultaneously in the ultrasound sonogram for a bistatic long-offset measurement shown in Figure 8c. The wave speeds were determined from the slope of each wave event. The shallow slope, or the faster wave, is that of the longitudinal wave (V_p), and the steeper slope is that of the shear wave (V_s). Once these speeds were determined, the elastic moduli were computed from the expressions shown in Figure 8 [25].

The results shown in Figures 7 and 8 are a first in the medical literature. We produced elastographic and measurements of bone defects by using skin- and eye-safe laser ultrasound on test tissue samples and without a surface treatment to enhance the optical return for a laser vibrometer measurement. In addition, it is likely a first that we were able to show that shear wave events in bone can be measured with a totally noncontact laser method.

Novel N-CLUS Applications

The N-CLUS concept was initiated and formulated for multiple medical applications that utilize anatomical imaging and elastographic capabilities separately or combined. There are currently no optoacoustic systems for medical applications in existence or in near-term development. The ability to perform contact-free ultrasound scans could improve the image quality and portability of ultrasound tomographic devices and could expand ultrasound applications. The main benefit of N-CLUS is that the noncontact optical approach attempts to minimize operator variability. Such an approach is envisioned to allow repeat-visit ultrasound measurements that can be compared to each other with enough accuracy to enable change detection. Recent work by other groups has focused on using fixed-frame water tank/gel bath

ROBERT W. HAUPT, CHARLES M. WYNN, MATTHEW R. JOHNSON, JONATHAN R. FINCKE, XIANG ZHANG, BRIAN W. ANTHONY, AND ANTHONY E. SAMIR

(a)

(b)

Scan 1   Scan 2   Scan 3   Scan 4

IN VIVO ELASTIC MODULI DETERMINED FROM ULTRASONIC VELOCITIES

YOUNG'S MODULUS	POISSON'S RATIO	BULK MODULUS	SHEAR MODULUS
$\rho V_{\mathrm{s}}^2 \left(3V_{\mathrm{P}}^2 - 2V_{\mathrm{s}}^2 \right)$ $/\left(V_{\mathrm{P}}^2 - V_{\mathrm{s}}^2 / 3 \right)$	$\left(V_{\mathrm{P}}^2 - 2V_{\mathrm{s}}^2 \right)$ $/2\left(V_{\mathrm{P}}^2 - V_{\mathrm{s}}^2 \right)$	$\rho\left(V_{\mathrm{P}}^2 - 4V_{\mathrm{s}}^2 / 3 \right)$	ρV_{s}^2
21.0 GPa	0.3474	22.5 GPa	7.6 GPa

where

Longitudinal (V_{P}) is 3890 m/s

Rayleigh (V_{s}) is 1711 m/s

Shear is Rayleigh / 0.91 = 1881 m/s

(c)

FIGURE 8. The N-CLUS system can be used to determine a bone sample's elastographic properties by using a far-offset receiver and source geometry to induce an ultrasonic surface wave in the bone (a). The wave traveled along the outer layers of the bone and returned to the surface, where it was measured by the laser Doppler vibrometer (b). The surface longitudinal and shear waves measured by N-CLUS are shown in the ultrasound sonogram created from this demonstration (c). The shear wave speed, compressional wave speed, and bone density are represented by V_{s}, V_{P}, and ρ, respectively. The wave speeds were calculated from the slope of each wave, and the elastic moduli were determined by plugging the wave speeds into mathematical equations (c).

approaches for performing change-detection ultrasound, such as tomography for cancer screening. While the images obtained with these systems have resolutions comparable to those of images obtained with MRI, these systems are generally designed exclusively for breast imaging, are unlikely to become portable, and cannot detect high-frequency shear waves (because the transducers are housed in water, shear attenuates drastically over submillimeter distances).

Applying ultrasound techniques to soft tissue containing bone presents additional imaging challenges. The strong refraction, attenuation, and scattering of transmitted acoustic waves used for imaging must be overcome. The ability to measure all waves (longitudinal,

shear, etc.) at many locations by using noncontact optoacoustic techniques promises to aid in meeting these challenges. Thus, a noncontact optoacoustic system capable of accurate characterization of bone and the surrounding soft tissue has the potential for unique applications, several of which we have explored: (1) improving prosthetic fittings by integrating internal tissue and bone structural information into the socket design process; (2) monitoring bone-density deterioration related to osteoporosis; (3) scanning load-bearing bones for stress fractures and early stages of damage; and (4) better quantifying the progression of neuromuscular disease.

N-CLUS Approach for the Detection and Imaging of Bone Injuries

The N-CLUS has the potential to detect and monitor common and debilitating bone injuries, such as stress fractures. When the bistatic far-offset acquisition geometry is utilized, N-CLUS generates refracted surface waves that produce the acoustic signatures of simulated bone defects that are similar in scale to stress fractures. In the demonstration shown in Figure 9, we introduced a 1-millimeter-deep by 1-millimeter-wide drill hole in the surface of a rib bone embedded in beef tissue. The beef tissue lying over the bone was approximately 2 to 4 millimeters thick. Once the small drill hole was made, the beef tissue was repositioned over the drill hole to maintain tissue continuity as would be expected in an actual tissue-bone setting.

Several scan lines were collected along the length of the bone. As observed in Figures 9b and 9c, refracted surface longitudinal and shear wave events dominate the sonogram. The green circled area shows the acoustic interference signature from these waves. It is important to note that though the simulated defect may be similar in scale to a common stress fracture, the material property contrasts may be more subtle in an actual stress fracture and result in a less pronounced signature, if any. Nonetheless, this example shows there may be potential in using the N-CLUS approach to detect, image, and monitor bone injuries such as stress fractures.

Prosthetic Fitting Using N-CLUS

We next show the potential of N-CLUS to acquire images of large-scale bone surrounded by soft tissue for applications in elastography, prosthetic limb fitting,

and prosthetic socket design. We employed a simulated residual limb (called a phantom model) containing a bone/muscle complex to demonstrate how N-CLUS's reflection and tomographic imaging capabilities could be used to optimize prosthetic fitting for an amputated limb.

The phantom limb shown in this example was constructed from a PVC pipe segment embedded in a copolymer soft tissue material. To reconstruct an image by using N-CLUS, one must know the target object's geometry, its surface topography, and the receiver and source locations. To obtain the topography, a NextEngine 3D laser scanned the phantom's surface and output the phantom's geometry in a 3D point cloud. A reference mark on the phantom allowed alignment between the laser source and receiver to the 3D-scan reference frame. Once the 3D scan was completed, the phantom was placed on a Sigma-Koki SGSP-YAW rotation stage with a 0.005-degree angular resolution. The N-CLUS source was driven by an optical parametric oscillator Q-switched laser with a fixed wavelength of 1550 nanometers, a 3-nanosecond pulse width, and a 3-millimeter projected spot diameter on the phantom. A Polytec OFV 505 laser vibrometer was used to measure the induced ultrasonic returns at a fixed location on the phantom's surface.

The data-acquisition process consisted of sending a laser pulse that excited a photoacoustic converted ultrasonic wave at the phantom's surface and receiving the return propagating wave with the laser vibrometer. For each transmit-and-receive location, the signal was averaged 64 times to obtain the SNR needed to create reasonable images and estimate travel times. Next, the stage rotated by two degrees, and the pulse-receive process was repeated until the stage completed a full 360-degree revolution. Once a full revolution was completed, the laser vibrometer's position was moved 30 degrees, and the pulse receiving process was repeated again. The data acquired during one full revolution of the rotation stage are equivalent to the data acquired by transmitting on one element and receiving on all others for a 180-element circular array fixed to the phantom's surface. In total, the dataset is analogous to sequentially transmitting on 12 elements each spaced by 30 degrees and receiving on all other elements for each transmit.

Standard migration routines produced the pulse echo image in Figure 10c, and the blockwise inversion technique [19] produced the sound-speed map in Figure 10d. One

ROBERT W. HAUPT, CHARLES M. WYNN, MATTHEW R. JOHNSON, JONATHAN R. FINCKE, XIANG ZHANG, BRIAN W. ANTHONY, AND ANTHONY E. SAMIR

No defect signature

Compressional wave

Ultrasound travel time (μsec)

Scan line 1   Scan line 2   Scan line 3   Scan line 4

Rayleigh shear wave

Cross range (cm)

(b)
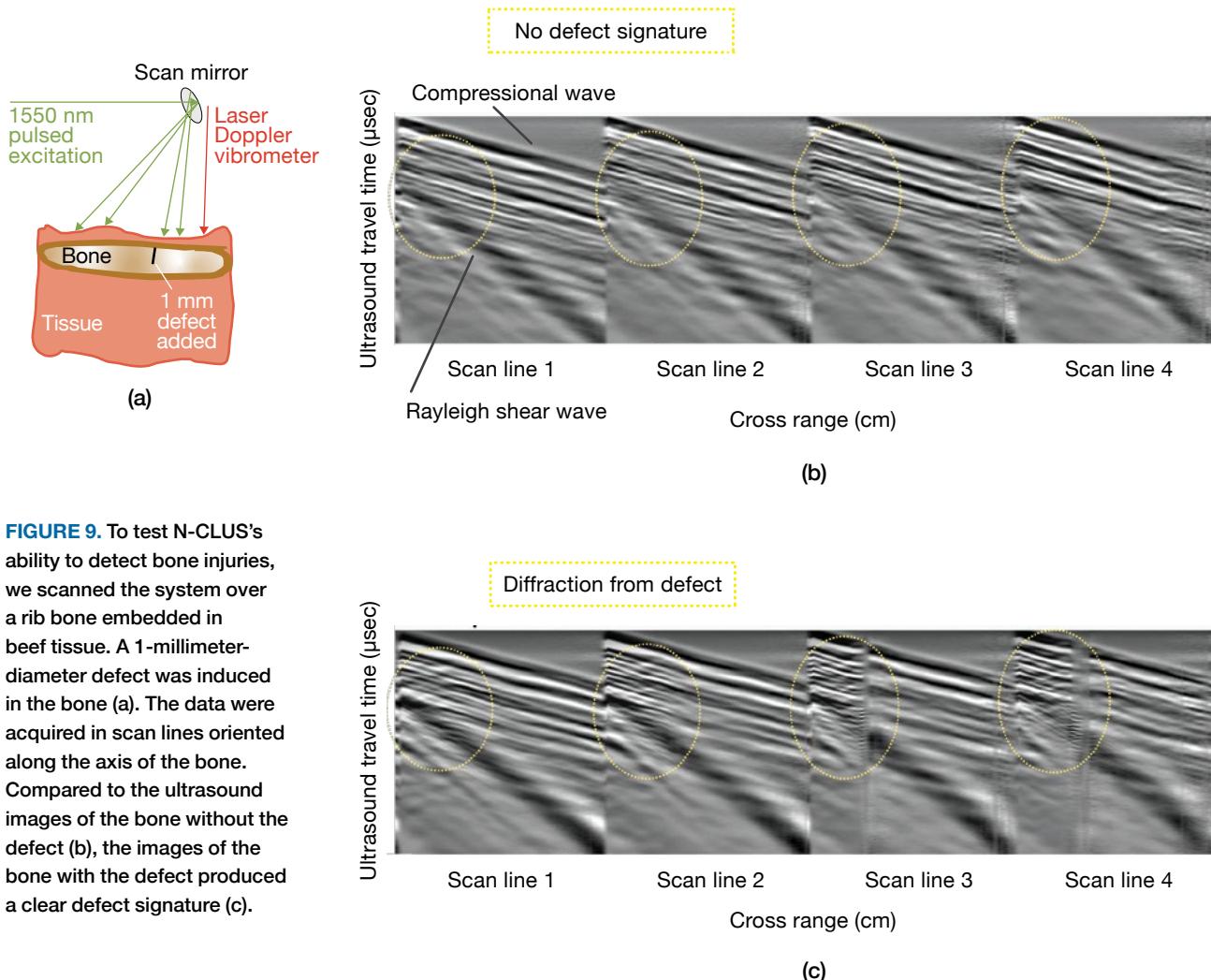
FIGURE 9. To test N-CLUS's ability to detect bone injuries, we scanned the system over a rib bone embedded in beef tissue. A 1-millimeter-diameter defect was induced in the bone (a). The data were acquired in scan lines oriented along the axis of the bone. Compared to the ultrasound images of the bone without the defect (b), the images of the bone with the defect produced a clear defect signature (c).

Scan mirror

1550 nm pulsed excitation

Laser Doppler vibrometer

Bone

Tissue

1 mm defect added

(a)

Diffraction from defect

Ultrasound travel time (μsec)

Scan line 1   Scan line 2   Scan line 3   Scan line 4

Cross range (cm)

(c)

of the critical issues with reconstructing images from laser ultrasound data is localizing the transmitter and receiver locations on the target. In this case, the use of a 3D scan and a rotational stage with a stationary target provided known measurement positions in a precise way, allowing the use of standard ultrasound reconstruction algorithms. The pulse echo image showed the diameter of the pipe to be 0.035 meters, and the true radius was 0.0334 meters. The estimate of the limb's sound speed was 1396 meters/second, and the estimate of the pipe's sound speed was 1934 meters/second. The true sound speed was 1406 meters/second for the limb and 2100 meters/second for the PVC pipe. As evidenced by the results, the system is able to quantify a tissue-like target with accuracy on the order of a few acoustic wavelengths.

Developing N-CLUS for Practical Medical Use

The demonstration system shown in Figure 2 asserts that N-CLUS can produce anatomical images and elastographic data for several applications. However, the N-CLUS system in its present form is not sufficiently operational for clinical use. Significant improvements need to be developed that would enable N-CLUS to provide images and data that are comparable and competitive with those of current state-of-the-art ultrasound systems used in medical practice. However, new components on the market now can enable a rack-mounted system that may be fast enough and sensitive enough to be used in a clinical or remote field-forward station in the very near term.

High-PRF Photoacoustic Source

High-power, high-PRF pulsed lasers that can produce photoacoustic ultrasound while operating within skin- and eye-safety limits have recently become available on the commercial market. We observed that a single optical pulse with a power of 1 millijoule and 1- to 10-nanosecond duration can easily generate a strong ultrasonic wave in tissue in the desired acoustic band of a few megahertz. These optical sources now can operate at a PRF of 5 kilohertz and provide continuous time coverage when the system listens for signals comprising an individual time-series trace of 200 microseconds. Current pulsed optical systems have been reduced in size and can be rack mounted. Private companies are

(a)

(b)

(c)

(d)
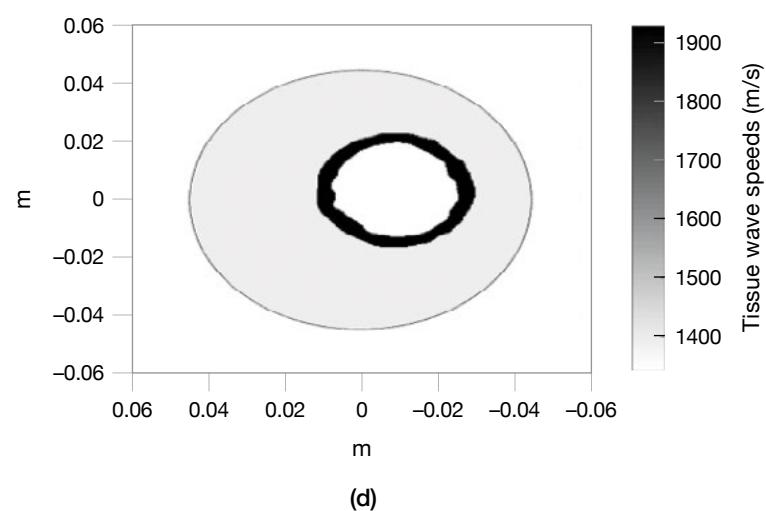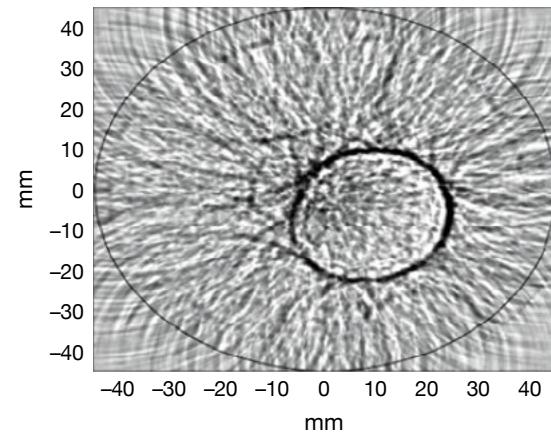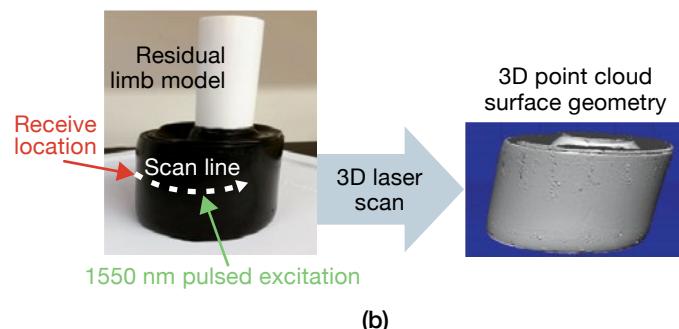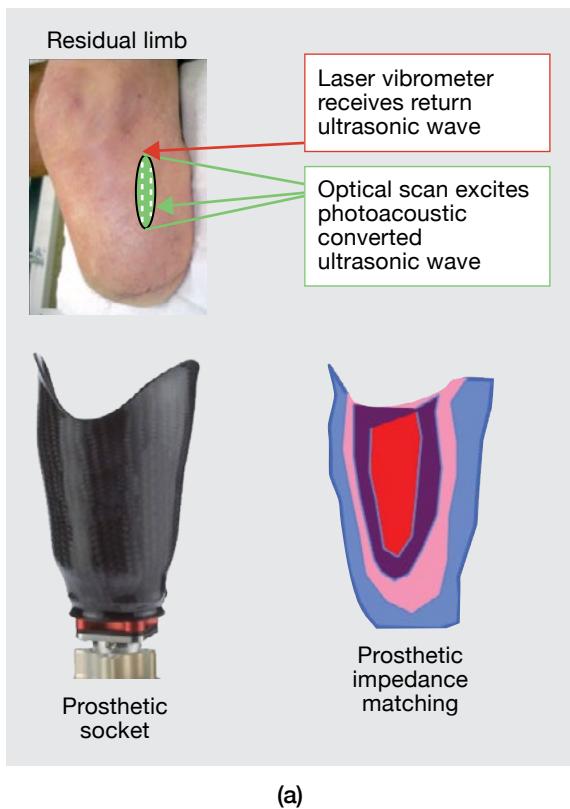
FIGURE 10. The N-CLUS system can potentially be used to image a residual limb to optimize prosthetic fitting for amputees (a). We constructed a simulated residual limb (phantom) out of a PVC pipe surrounded by a soft tissue–like material (b). The phantom was first scanned by a 3D laser and then placed on a rotating platform before it was scanned by the N-CLUS system, which was held at a constant location. A pulse echo image (c) and a sound speed map (d) of the phantom were produced from the scans. In the pulse echo image, the PVC pipe (which simulates bone) was distinct from the surrounding soft tissue material. The speed of the ultrasonic wave traveling in the soft tissue material and in the PVC pipe was determined from the N-CLUS measurements. These speeds closely matched the same wave speeds that were determined from measurements produced by contact transducers.

pushing to miniaturize these sources, and handheld, high-power, high-PRF systems are anticipated to be available in a year or two. The improved PRF can decrease N-CLUS's image data-acquisition time from 10 minutes to around 1 second. To accomplish the high PRF while maintaining skin and eye safety, we constructed a 1950-nanometer-wavelength Q-switched laser that generates optical 2-millijoule, 5-nanosecond pulses and 1000 to 5000 of these pulses per second. A simple fast scanning mirror can be used to steer the optical excitation beam on the skin's surface.

Multipixel Laser Vibrometer

In our studies, the N-CLUS demonstration system's single-pixel laser vibrometer operating at 1550-nanometer and 10-milliwatt power (skin and eye safe) was able to measure returning ultrasonic signals. However, variations in optical speckle and in the surface diffusion of tissue caused significant subsecond time-scale fluctuations in the noise floor and in the performance of the laser vibrometer. To overcome these fluctuations, the laser measurement required averaging consecutive time-series traces at a given location to reduce the noise variance enough that the ultrasonic signal of interest could be observed above the noise floor. The acquisition time was increased even further because producing the N-CLUS image involved averaging 64 data trace collections per scan position location.

To implement a laser vibrometer ultrasound receiver that can meet the area coverage rate requirements in a clinical setting, a multipixel laser vibrometer system will be necessary. The multipixel configuration should be designed to capture multiple speckle lobes in the receive laser beam simultaneously for a given station position. Statistically, some of the speckle lobes will capture small noise contributions, while other lobes will capture large contributions. The pixels associated with large noise can be excluded, while pixels with small noise will be kept and averaged to reduce the noise variance further.

Currently, commercial vendors have been developing multipixel vibrometer systems that can measure ultrasonic vibrations on surfaces. These systems, which we have incorporated into the N-CLUS receiver design, can be readily adapted to use skin- and eye-safe optical wavelengths and powers for medical ultrasound imaging systems. To accomplish simultaneous averaging, we constructed a 48-pixel skin- and eye-safe LDV, pictured in Figure 11. This vibrometer measures several speckle lobes of the vibrational ultrasound signal simultaneously, and we keep the strong SNR returns, discard the weak returns, and average.
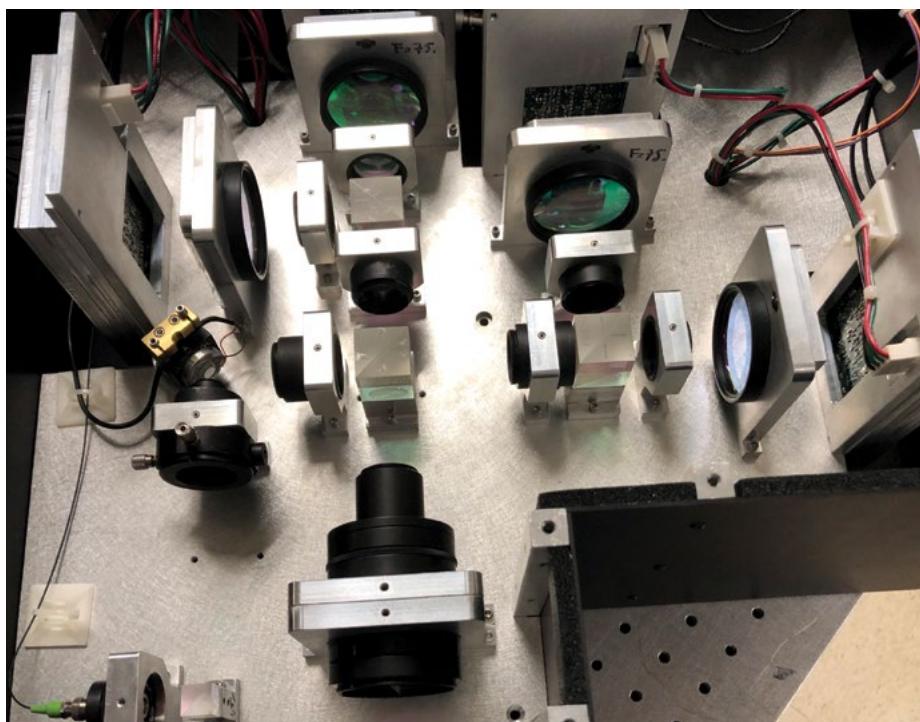
FIGURE 11. The skin- and eye-safe laser Doppler vibrometer utilizes 48 pixels simultaneously. The system has a bandwidth of 5 megahertz and uses 2-inch collection optics that can be operated from a standoff distance of 100 to 500 millimeters. The laser has an output of 10 milliwatts and an adjustable manual focus.

Stand-Mounted and Handheld N-CLUS System

Results from our studies show that a rack-mounted optoacoustic system for medical imaging could be assembled in the near term. As demonstrated in Figure 12, the optical head would be small enough to mount on a stand, and the head could be positioned over the target area on the patient's body and then scanned to acquire ultrasound data. Such a system could be used in a clinical setting or remote field-forward station. A high-PRF photoacoustic source and multipixel laser vibrometer can be implemented in the system.

In the long term, there is a significant desire to develop a small, portable, handheld noncontact optoacoustic system such as the one envisioned in Figure 13. Such a system will require miniaturization of components and will also require methods to compensate for operator motion and look angles, as is necessary for conventional handheld contact ultrasound systems. The use of chip-scale lidar technology is proposed to enable miniaturization and compensate for motion. Large numbers of pixels are being proposed to meet operational coverage rates required for full-body-scan ultrasound. At Lincoln Laboratory, we are investigating the components

needed for this capability. Technology such as Lincoln Laboratory–developed coherent focal plane arrays are being considered as a solution to handle the high data rates and coverage of medical ultrasound at frequency bandwidths of several megahertz. ∎
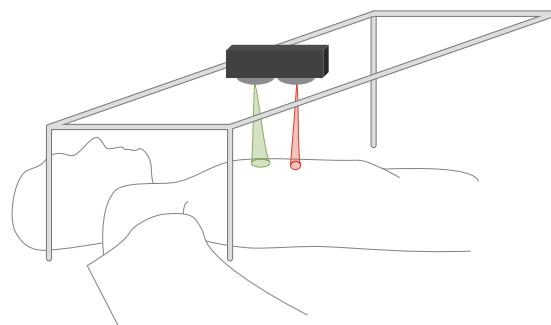
FIGURE 12. A simple stand mount for acquiring static N-CLUS measurements is proposed. The optical head is fixed in the stand and can be positioned anywhere in the frame by using a joy stick. Optical excitation and measurement laser beams can be scanned over a region of the body by using fast scanning mirrors. The electronics associated with the optical head can be rack mounted. The system would be portable enough to be assembled and used in a clinical setting or in a remote field-forward station.

FIGURE 13. A portable handheld N-CLUS system concept is illustrated (a). The system would contain a photoacoustic source and a laser Doppler vibrometer on the same unit. Coherent focal plane arrays (b) are being considered as an element to integrate into medical ultrasound, but significant development is needed to implement chip-scale lidar and coherent focal plane array technology on medical devices.

ROBERT W. HAUPT, CHARLES M. WYNN, MATTHEW R. JOHNSON, JONATHAN R. FINCKE, XIANG ZHANG, BRIAN W. ANTHONY, AND ANTHONY E. SAMIR

References

1. W. Hendee and E. Ritenour, *Medical Imaging Physics*. New York: Wiley-Liss, 2002.

2. American Thyroid Association, "American Thyroid Association Guidelines," 2015, available at https://www.thyroid.org/professionals/ata-professional-guidelines/.

3. E.A. Eisenhauer, P. Therasse, J. Bogaerts, L.H. Schwartz, D. Sargent, R. Ford, et al., "New Response Evaluation Criteria in Solid Tumors: Revised RECIST Guideline (Version 1.1)," *European Journal of Cancer*, vol. 45, no. 2, 2009, pp. 228–247.

4. R.W. Haupt, C.M. Wynn, K. Holman, J. Kyung, and A. Samir, "Noncontact Laser Ultrasound (NCLUS)—Method to Mitigate Operator and Inter-operator Variability," U.S. patent application no. 62/309497, applied 17 Mar. 2016.

5. C.A. Pulafito, R. Birngruber, T.F. Deutsch, J.G. Fujimoto, D. Stern, and B. Zysset, "Laser-Tissue Interactions in the Nanosecond, Picosecond, and Femotosecond Time Domains," *Photoacoustics and Photothermal Phenomen II—Springer Series in Optical Sciences*, vol. 62, 1990, pp. 420–427.

6. L.V. Wang, "Tutorial on Photoacoustic Microscopy and Computed Tomography," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 14, no. 1, 2008, pp. 171–179.

7. M. Xu and L. Wang, "Photoacoustic Imaging in Biomedicine," *Review of Scientific Instruments*, vol. 77, no. 4, 2006.

8. C. Li and L.V. Wang, "Photoacoustic Tomography and Sensing in Biomedicine," *Physics in Medicine and Biology*, vol. 54, no. 19, 2009, pp. R59–R97.

9. B. Yin, D. Xing, Y. Wang, Y. Zeng, Y. Tan, and Q. Chen, "Fast Photoacoustic Imaging System Based on 320-Element Linear Transducer Array," *Physics in Medicine and Biology*, vol. 49, no. 7, 2004, pp. 1339–1346.

10. G. Xu, C. Wang, T. Feng, D. Oliver, and X. Wang, "Noncontact Photoacoustic Tomography with a Laser Doppler Vibrometer," *Proceedings of the Society of Photo-optical Instrumentation Engineers (SPIE)*, vol. 8943, 2014.

11. G. Rousseau, A. Blouin, and J. Monchalin, "Noncontact Photoacoustic Tomography and Ultrasonography for Tissue Imaging," *Biomedical Optics Express*, vol. 3, no. 1, 2012, pp. 16–25.

12. R.W. Haupt and C.M. Wynn, "System and Method for Noncontact Ultrasound," U.S. patent application no. 14/538698, issued 15 Jun. 2019.

13. A.R. Fisher, A.J. Schissler, and J.C. Schotland, "Photoacoustic Effect for Multiply Scattered Light," *Physical Review E*, vol. 76, no. 3, 2007.

14. J. Ripoll and V. Ntziachristos, "Quantitative Point Source Photoacoustic Inversion Formulas for Scattering and Absorbing Media," *Physical Review E*, vol. 71, no. 3, 2005.

15. D.A. Schurig, G. Klunder, M. Shannon, R. Russo, and R. Silva, "Signal Analysis of Transients in Pulsed Photoacoustic Spectroscopy," *Review of Scientific Instruments*, vol. 64, no. 2, 1993, pp. 363–373.

16. L.A. Jiang, M. Albota, R. Haupt, J. Chen, and R. Marino, "Laser Vibrometry from a Moving Ground Vehicle," *Applied Optics*, vol. 50, no. 15, 2011, pp. 2263–2273.

17. W.M. Telford, L.P. Geldart, R.E. Sheriff, and D.A. Keys, *Applied Geophysics*. Cambridge, U.K.: Cambridge University, 1990.

18. Laser Institute of America, "ANSI Z136.1—2007," 2007, available at https://www.lia.org/store/product/ansi-z1361-2014-safe-use-lasers-electronic-version.

19. X. Zhang, J.R. Fincke, and B.W. Anthony, "Single Element Ultrasonic Imaging of Limb Geometry: An In-Vivo Study with Comparison to MRI," paper in *Proceedings of the Society of Photo-optical Instrumentation Engineers (SPIE)*, vol. 9790, 2016.

20. R. Haupt, X. Zhang, J. Fincke, J. Richardson, J. Hughes, B. Anthony, and A. Samir, "Non-Contact Laser Ultrasound System for Medical Imaging and Elastography," paper in *Proceedings of IEEE International Ultrasonics Symposium*, 2019, pp. 1921–1925.

21. X. Zhang, J.R. Fincke, C.M. Wynn, M.R. Johnson, R.W. Haupt, and B.W. Anthony, "Full Noncontact Laser Ultrasound: First Human Data," *Nature—Light: Science and Applications*, vol. 8, no. 119, 2019.

22. R. Haupt and A. Samir, "Non-Contact Laser Ultrasound for Biomedical Applications," paper presented at American Institute of Ultrasound in Medicine (AIUM), Orlando, May 2019.

23. J. Fincke, R. Haupt, C. Wynn, X. Zhang, D. Rivera, and B. Anthony, "Characterization of Laser Ultrasound Source Signals in Biological Tissues," *Journal of Biomedical Optics*, vol. 24, no. 2, 2018, pp. 1–11.

24. R. Haupt, C. Wynn, J. Fincke, X. Zhang, B. Anthony, and A. Samir, "Non-Contact Laser Ultrasound for Medical Imaging and Diagnosis," paper in *Proceedings of the IEEE International Ultrasonics Symposium*, 2017.

25. R. Sheriff, *Encyclopedic Dictionary of Exploration Geophysics*. Tulsa: Society of Exploration Geophysicists, 1999.

Photoacoustic Phenomenology and Ultrasonic Wave Propagation

Optical Absorption Process

In the first stage of the photoacoustic process, photons are absorbed by particles composing a tissue volume, where the absorption coefficient

$$\mu_a = \rho \, \sigma_a \text{ where }, \sigma_a = -4\frac{2\pi a}{\lambda}\pi a^2 Im\left\{\frac{n_1 - n_0}{n_1 + 2n_0}\right\}$$

and ρ and σ_a are the particle density and cross-sectional area; a is the particle radius, where a is orders of magnitude less than the optical wavelength; and n_1 and n_0 are the refraction indices of the absorbing material and an infinite homogeneous nonabsorbing medium.

For an optical pulse incident on tissue particles, the total absorbed energy, E_a, becomes

$$E_a(\mathbf{r},t) = \mu_a \int_{4\pi} I(\mathbf{r},t,\hat{\mathbf{s}}) \, d\Omega = \mu_a U^{inc}(\mathbf{r},t)$$

where I is the specific intensity absorbed by the tissue particles at a position $\mathbf{r}$ from light incident in a direction $\hat{\mathbf{s}}$. U^{inc} is the average incident intensity with units of joules/square centimeter (J/cm^2). The governing relationship establishing tissue deformation, and thus acoustic or elastic wave generation, evolves from the tissue temperature increase caused by the absorbed energy where

$$\rho_m C \frac{\partial T(\mathbf{r},t)}{\partial t} - \kappa \nabla^2 T(\mathbf{r},t) = E_a(\mathbf{r},t)$$

and ρ_m, C, κ, T are the tissue mass density, specific heat, thermal conductivity, and temperature, respectively. The first term in the above equation describes the temperature increase due to optical absorption and diffusion. The optical diffusion is several orders of magnitude larger than the thermal diffusion; thus, the second term in the equation is negligible.

Optical Propagation Process

The next critical component to understanding the process of photoacoustic phenomenology is the effect of optical propagation into a scattering media such as complex biological tissue. The materials composing a tissue mass are considerably heterogeneous. Blood hemoglobin, for example, is highly absorptive, while other tissue cells are highly reflective for wavelengths on the order of 800 nanometers (a commonly used optical wavelength in photoacoustic tomography). However, optical wavelengths greater than 1000 nanometers are strongly absorbed and less sensitive to tissue property variations. The average incident energy can be derived by showing the relationship between the incident energy density in the time domain, where

$$U^{inc}(\mathbf{r_s},\mathbf{r},t) = \frac{S_0}{(4\pi Dct)^{3/2}} \exp\left[\frac{|\mathbf{r_s} - \mathbf{r}|}{4Dct} - \mu_a |\mathbf{r_s} - \mathbf{r}|\right]$$

Optical Conversion to Ultrasound Wave Propagation

Explained next is the last critical component in describing the photoacoustic conversion of light to pressure and the resultant acoustic wave propagation. It is the acoustic or elastic wave that can be measured by an optical receiver such as a laser Doppler vibrometer or conventional contact transducer. For simplicity, we consider the case of an inviscid fluid to demonstrate the generation and propagation of the longitudinal or compressional wave from incident light. We start with the linear force equation

$$\rho_m \frac{\partial^2 \mathbf{u}(\mathbf{r},t)}{\partial t^2} = -\nabla p(\mathbf{r},t)$$

where $\mathbf{u}$ is the acoustic displacement, and p is the acoustic pressure. The relationship between the heat source and the

ROBERT W. HAUPT, CHARLES M. WYNN, MATTHEW R. JOHNSON, JONATHAN R. FINCKE, XIANG ZHANG, BRIAN W. ANTHONY, AND ANTHONY E. SAMIR

resultant pressure is shown below in terms of the optical average intensity and optical absorption coefficient:

$$\nabla^2 p(\mathbf{r},t) - \frac{1}{v_s^2}\frac{\partial^2 p(\mathbf{r},t)}{\partial t^2} = \frac{\beta}{C}\Big[\mu_a + \nabla\mu_a(r)\Big]\frac{\partial U(\mathbf{r},t)}{\partial t}$$

The pressure distribution along the tissue column resolves to

$$p(\mathbf{r},t) = p_0(\mathbf{r},t) + \frac{\beta}{4\pi C}\int_V \frac{\mathbf{dr}'}{|r-r'|}\Delta\mu_a(\mathbf{r}')\times$$

$$\left.\left|\frac{\partial U^{\text{inc}}(\mathbf{r}',t')}{\partial t'}\right|\right|_{t'=t-|\mathbf{r}-\mathbf{r}'|/v_s}$$

where $p_0(\mathbf{r}, t)$ is the incident pressure at the onset of the tissue column

$$p_0(\mathbf{r},t) = \frac{\beta\mu_a}{4\pi C}\int_V \frac{\mathbf{dr}'}{|r-r'|}\left.\left|\frac{\partial U^{\text{inc}}(\mathbf{r}',t')}{\partial t'}\right|\right|_{t'=t-|\mathbf{r}-\mathbf{r}'|/v_s}$$

About the Authors

Robert W. Haupt is a senior technical staff member in the Active Optical Systems Group at Lincoln Laboratory. His research emphasis is on coherent laser radar for imaging vibration and acoustic fields and for determining the material properties of biological tissue from a standoff distance. He has developed new system concepts, including the noncontact laser ultrasound for medical imaging, photoacoustic sensing of explosives, buried landmine detection, and seismic cloaking. He is the principal investigator (along with Charles Wynn) and the program manager of the N-CLUS project, which is now funded by the U.S. Army. He has a master's degree in geophysics from the Pennsylvania State University and a master's degree in mechanical engineering from Thayer School of Engineering at Dartmouth College in Hanover, New Hampshire. He has a bachelor's degree in physics and a bachelor's degree in meteorology from the State University of New York at Oswego.

Charles M. Wynn is a technical staff member in the Advanced Capabilities and Technologies Group at Lincoln Laboratory. He is currently leading several photoacoustics efforts and a hyperspectral imaging program. Much of his research in recent years has been on the laser-based detection of trace explosives. He is also engaged in research using photoacoustic spectroscopy for chemical detection and has spearheaded a biotechnology initiative that leverages a novel method for photoacoustic sensing. Most recently, he developed an ultra-compact hyperspectral imager by using computational imaging techniques. He is currently using photoacoustics to create useful audio that can be listened to from a distance. He earned a bachelor's degree in physics from the University of Connecticut, a master's degree in physics from Carnegie Mellon University, and a doctoral degree in physics from Clark University.

Matthew R. Johnson is an associate staff member in the Human Health and Performance Systems Group at Lincoln Laboratory. His work centers on the development of technology for national security and health care, with a specific focus on applying defense technology to the health care sector. He completed his undergraduate studies in electrical engineering at Colorado State University and has completed his MS degree in electrical engineering at MIT.

Jonathan R. Fincke is a research scientist at Phillips Ultrasound in the Boston area and has engaged in research for the development of systems and algorithms for noncontact quantitative ultrasound imaging of soft tissue and bone. He received a bachelor's degree in mechanical engineering from the University of New Hampshire, a master's degree in oceanographic engineering from the MIT/Woods Hole Oceanographic Institution Joint Program, and a doctoral degree in mechanical engineering from MIT.
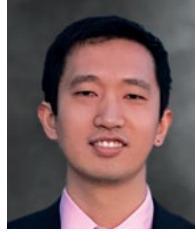
Xiang Zhang is a postdoctoral scientist in the Department of Mechanical Engineering at MIT. He has engaged in research focusing on the design of electromechanical systems to enhance medical ultrasound, 3D tomographic ultrasound for tissue property characterization, and noncontact laser ultrasound. He received a bachelor's degree in mechanical engineering from the University of Maryland–College Park. During his undergraduate studies, he completed a co-op at the National Institute of Standards and Technology. He holds master's and doctoral degrees from MIT.

Brian W. Anthony is a principal research scientist at the Department of Mechanical Engineering and the Institute for Medical Engineering and Science at MIT. Currently, he is a co-director of the Medical Electronic Device Realization Center and the director of the MIT Master of Engineering in Manufacturing program. His research includes the development of instrumentation and measurement solutions for manufacturing systems, medical diagnostics, and imaging systems.

Anthony E. Samir is an abdominal and interventional radiologist and sonologist at Massachusetts General Hospital. He has a background in medical device innovation. His research interest is in translational applications and technical improvements in ultrasound with a focus on shear wave elastography and tissue characterization. He has expertise in the development, validation, and translation of quantitative imaging biomarkers with ultrasound. He has led numerous projects in liver imaging and in breast imaging. He received an MD degree from University of the Witwatersrand, Johannesburg, Gauteng, and a master's degree in public health from the Harvard School of Public Health.

Biomechanical Sensing and Algorithms

Joseph J. Lacirignola, James R. Williamson, Shakti K. Davis, and Whitney T. Young

Lincoln Laboratory has been developing software and hardware systems that can help determine if a soldier's performance has degraded due to fatigue, injury, or another detrimental physical state. Biomechanical data collected outdoors with commercial sensing devices have shown that several algorithms and models could interpret body movements to deduce a soldier's gait asymmetries, fatigue level, and amount of load carriage. Custom shoe inserts have been developed to measure lower-leg forces and movements in outdoor environments. The ability to acquire and interpret lower-leg measurements outside of a laboratory without compromising the natural rhythms of a person's gait facilitates the monitoring of complex biomechanical actions in military environments. Collecting lower-leg measurements outdoors can help researchers develop new models and methods for monitoring gait changes and improving military training to reduce the rate of musculoskeletal injuries.

» **Lower-limb musculoskeletal injuries** (MSIs) are pervasive in military populations, with hundreds of thousands of MSIs reported every year [1]. These injuries include acute traumas (e.g., anterior cruciate ligament tears) and overuse injuries (e.g., stress fractures and medial tibial stress syndrome) [2, 3]. MSIs are a result of external stimuli and stressors, such as carrying a heavy load, that put strain on the complex biomechanical network of the lower legs. This network involves the interactions between muscles, joints, and the nervous system, and the dynamic forces and motions that result from such interactions. Training causes increased amounts of stress to load-bearing bones and joints because heavy load carriage and aggressive movements exert higher-than-normal forces on leg joints (hip, knee, and ankle) and bones (femur, fibula, tibia, and foot bones). Accurate measurements of the lower-leg forces and motions could inform decisions about how load is applied to each soldier during training and combat. These measurements could help prevent training injuries and lead to the development of training procedures that focus on maintaining a soldier's mobility and agility. Determining the dynamic forces and motions of the lower-limb movements and their interactions requires measuring lower-leg forces, acceleration, and displacements and using algorithms to elucidate the association between the different factors of lower-leg movement.

A variety of biomechanical sensors can be used to monitor leg movement, as shown in Figure 1. Measuring lower-leg forces and motions in a robust, repeatable
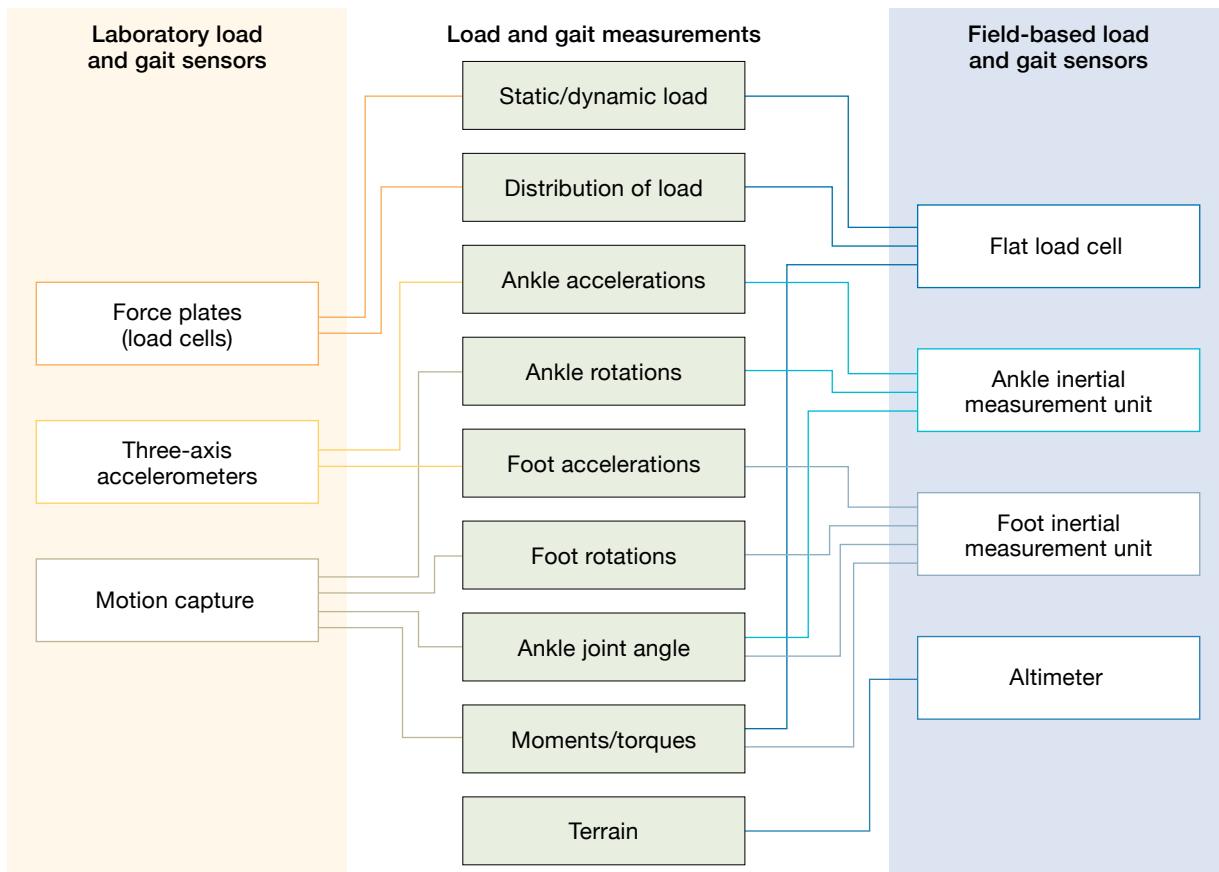
FIGURE 1. A comparison of laboratory- and field-based biomechanical measurement systems is illustrated. The far left column is a list of gold-standard laboratory load and gait sensors while the far right column is a list of field-based load and gait sensors. The middle column is a list of load and gait measurements. Lines from the laboratory and field sensors indicate which measurements can be optioned from each class of sensor.

manner is challenging to achieve outside the laboratory because existing biomechanical sensors are limited by factors such as a lack of accuracy or a design that impedes natural movement. Lower-leg kinematic gait models exist, but interpreting lower-leg measurements and turning them into actionable information is difficult because mobility, agility, or injury models do not exist.

In a laboratory setting, measuring lower-leg biomechanics is well established. Force plates and force-sensing treadmills measure ground reaction forces (GRFs), or the forces exerted by the ground on a body in contact with it. Fixed, multicamera motion-capture systems and accelerometers can directly quantify the kinematic and kinetic responses of lower-limb movement for static and dynamic conditions [4]. Although these laboratory-grade sensors are accurate and are considered the gold standard for biomechanical measurements,

they are cumbersome and difficult to translate to a field environment; thus, there is a lack of field data and a limited ability to determine the real-time impact of environmental stressors, such as different terrain types and sudden changes in elevation, on an individual's mobility, agility, or susceptibility to injury. There is also a lack of persistent, long-duration measurements that explicate what role repetitive motion or fatigue plays in injury or reduction in agility and mobility. In addition, little is known about how lower-limb movements, forces, and accelerations change during unloaded or loaded activity prior to and after an MSI is incurred.

Advances in electronics miniaturization have enabled some conventional laboratory measurement methods, such as motion tracking, motion capture, and load or pressure sensing, to be incorporated into wearable devices. Taking these biomechanical measurements

outside of the laboratory is a big step toward gathering much-needed real-world data. However, several of the developed devices still fall short compared to their gold-standard laboratory counterparts. Commercial wrist, head, limb, or torso-based motion-tracking devices (e.g., Fitbit and heart rate monitors) have unobtrusive forms but lack the sampling rates and accuracy needed for comprehensive biomechanical analysis. Wearable motion-capture systems provide more accurate measurements than motion-tracking devices and are often accompanied by biomechanical models to relate the sensors to each other during movement. But the improvement in accuracy comes at the price of obtrusive design components (e.g., multiple long wires or battery packs), insufficient battery life (one to two hours), and susceptibility to magnetic interference.

Force-sensing shoes and shoe inserts can measure GRFs outside the laboratory. Sensors developed for these devices currently lack robustness, dynamic range, accuracy, and forms that do not alter gait. Additionally, this group of sensors does not yet have analytic packages that can detect gait anomalies or state changes in lower-leg movement.

Several commercially available accelerometers or inertial measurement units (IMUs) are small and wearable, and can support field-based measurements. Shoe- and chest-worn accelerometers can be used to determine basic gait features and estimate the weight carried by an individual. However, further research and development are needed to mature the state of models and algorithms for these types of sensors and to convert the raw IMU data into actionable biomechanical information.

Estimating Gait Asymmetry from Commercial Shoe-Mounted Accelerometers

Gait asymmetry (nonidentical spatiotemporal parameters between limbs) can be a useful indicator of medical and pathological conditions, including MSIs, neurological damage associated with stroke or head trauma, and a variety of age-related disorders [11]. The ability of body-worn accelerometers to monitor and estimate gait asymmetry in real time can help researchers be aware of any medical conditions in a subject and administer timely interventions. At Lincoln Laboratory, accelerometers mounted on shoes have been used to measure an individual's foot motion during walking to estimate the

level of gait asymmetry. These estimates were achieved by computing simultaneous acceleration differences between the two feet to test the system's ability to reliably detect small asymmetries in the presence of natural variability in walking speeds and conditions. To test a gait asymmetry estimation algorithm, we conducted an experiment in which ankle weights were used to create asymmetries in the natural walking gait. The algorithm was able to estimate the direction and magnitude of gait asymmetries during continuous gait monitoring.

Gait Asymmetry Features

Several types of gait features have been investigated to assess gait asymmetry: stride timing (swing or stance times) [5–8], acceleration magnitude [9], and acceleration time-varying patterns and the degree of similarity between given peaks (autocorrelation) [9, 10]. These features have primarily been assessed in the context of clinical or laboratory measurements compared against normative values. While many studies assume gait symmetry for normative values, even a healthy individual may not have a symmetric gait [11]. Establishing the baseline gait measurements of a healthy individual and obtaining individualized gait characteristics are useful in evaluating changes in gait over time. Individualized gait asymmetry detection is also important because there is natural variability between different subjects' gaits.

The Lincoln Laboratory gait asymmetry estimation algorithm uses novel statistical features that characterize asymmetry in acceleration magnitudes and dynamical patterns. These features are combined in an individualized algorithm that estimates gait asymmetry relative to the user's baseline gait measurements in variable contexts, such as different walking surfaces and speeds.

Statistical foot-motion features are computed from the acceleration magnitudes of the left and right feet within fixed 20-second time frames. These features, which do not rely on the detection of stride events, use simple computations that are easy to implement locally on the foot sensor. Two types of motion features, magnitude and pattern features, are computed at the sensors and then combined across the two feet to yield asymmetry features. Magnitude asymmetry is computed using the standard deviations of the left and right feet's
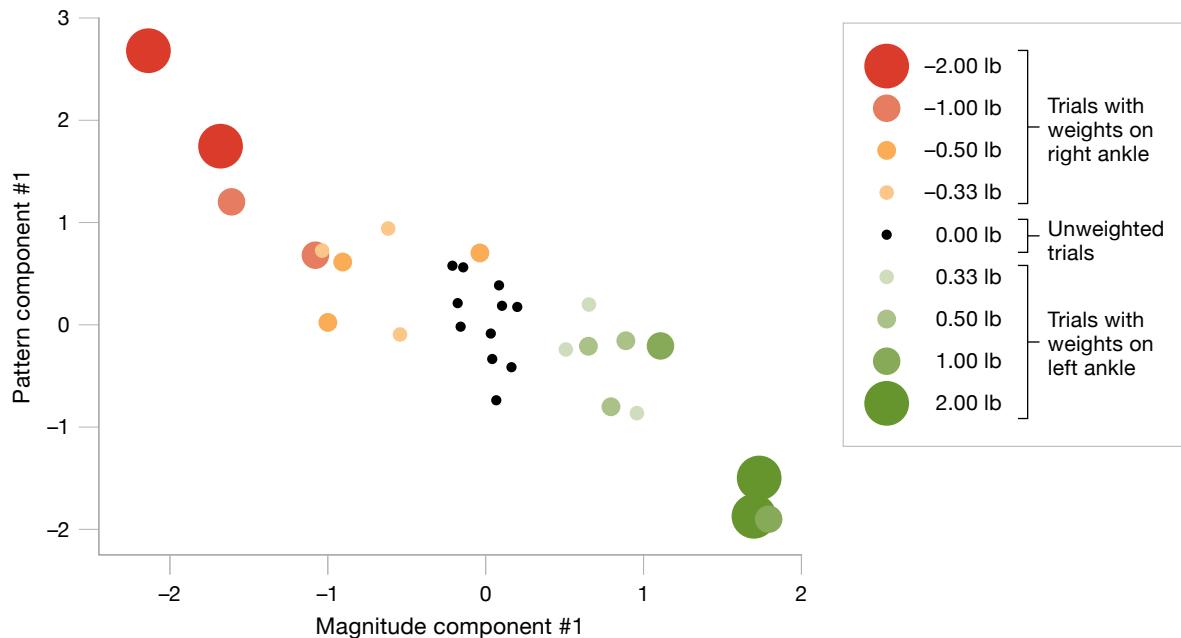
FIGURE 2. The mean feature values from 31 trials for Subject 6 are plotted above. On the *x*-axis is the magnitude feature, and on the *y*-axis is the first frame-based pattern principal component. Black circles indicate unweighted trials, shades of red indicate trials with weight on the right ankle, and shades of green indicate trials with weight on the left ankle. Marker size and shade indicate the size of ankle weights, which ranged from 0.33 to two pounds. The greater the induced weight of a trial (larger circles), the more the magnitude and pattern features deviate from those of the unweighted trials (small, black circles), indicating that gait asymmetry can be estimated from comparing the observed features.

magnitudes. Let x and y represent the standard deviation of the left and right feet's acceleration magnitude signals, computed over the fixed time frame. The magnitude asymmetry is computed by using Equation (1), the symmetry index function [12].

$$SI(x,y) = \frac{x - y}{0.5(x + y)} \qquad (1)$$

Pattern features register differences in the temporal stride dynamics between the two feet after magnitude differences have been factored out. Within-foot stride dynamics are captured by using autocorrelations of the acceleration magnitude signals. Autocorrelation peaks have been used by researchers to characterize gait from a single accelerometer [13]. In the Lincoln Laboratory algorithm, the autocorrelation function is regularly sampled out to the first peak, thereby representing time-delay correlation coefficients spanning a single stride. Differences between the autocorrelations of the two feet over these time delays constitute a raw pattern asymmetry feature vector. The first four principal components of this vector produce the final pattern asymmetry features.

Estimating and Detecting Asymmetry

The gait pattern for a healthy, uninjured individual was characterized on the basis of a normal distribution obtained from training set features from the same individual with no induced asymmetry. The induced gait asymmetry was then detected by measuring the distance of gait features from the normal gait feature distribution. To induce gait asymmetries in a controlled manner for 24 subjects, we added a range of weights to each subject's right or left ankle and obtained walking trial measurements. Figure 2 shows a scatter plot of the mean feature values computed from 31 separate indoor and outdoor walking trials for one subject. The scatter plot represents the magnitude asymmetry feature (symmetry index of standard deviation of acceleration magnitude) and the first pattern asymmetry feature (first principal component extracted from the autocorrelation differences). The unweighted trials are plotted with black circles. Right ankle-weighted trials are plotted with red circles and left ankle-weighted trials with green circles. The size of the circle indicates the magnitude of the weight. The magnitude and pattern features depend on which foot the load is applied to and the level of induced gait asymmetry.
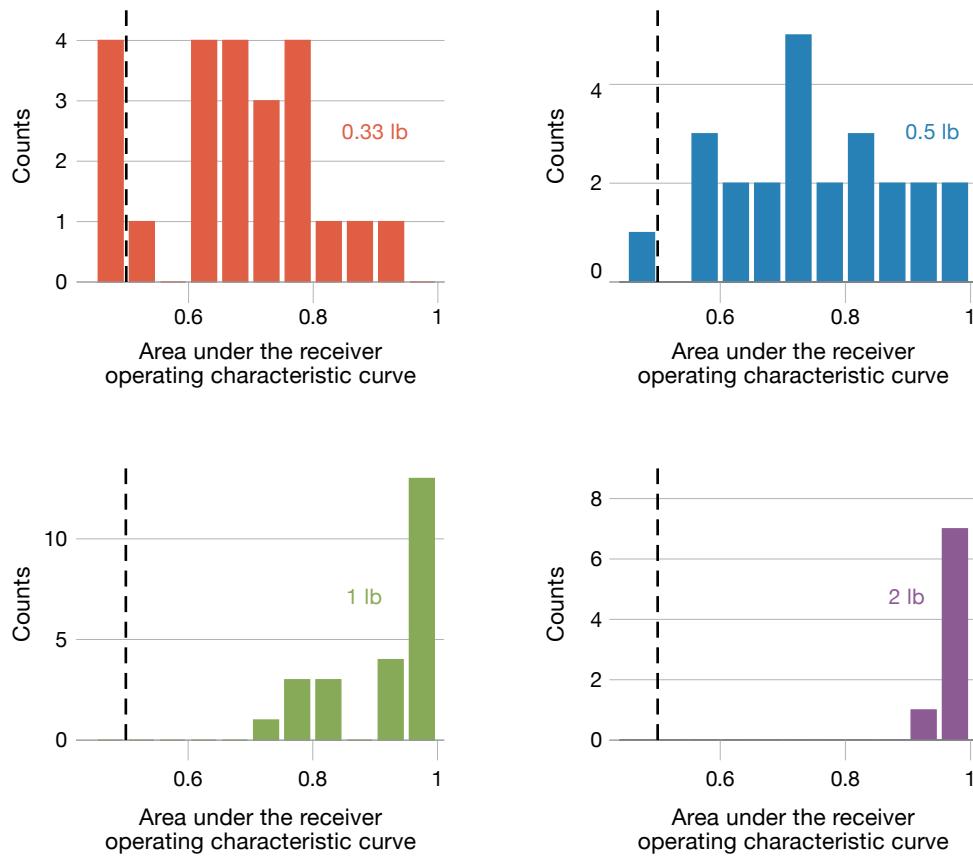
FIGURE 3. This figure shows histograms of gait asymmetry detection results across subjects in the four weighted conditions. Under each weight condition, an AUC (area under the receiver operating characteristic curve) value for each subject was computed that quantified the ability to detect the presence of an induced asymmetry in a trial when it is compared to unweighted control trials for the same subject. These histograms show the counts, across all the subjects, of different ranges of AUC values computed. AUC = 0.5 indicates chance-level accuracy and AUC = 1.0 indicates perfect accuracy. The trials with larger induced gait asymmetry had AUC distributions that were closer to 1.0, which means that these trials had a higher accuracy rate of asymmetry detection.

The trend for a single subject shown in Figure 2 holds across many subjects. Figure 3 summarizes the ability to detect, based on combined magnitude and pattern features, asymmetrical walking trials by using outlier detection. The outlier detection was performed by using subject-dependent background models constructed from three randomly selected unweighted walking trials. The distributions of an accuracy statistic, the area under the receiver operating characteristic curve (AUC), are plotted for four induced weight conditions. Larger induced gait asymmetries led to easier outlier detection.

Case Study: Use of Commercial Sensors to Track Recovery from Injury

The results on induced asymmetries raised the question, "How does the algorithm perform on gait asymmetries that arise from an actual overuse injury?" We addressed this question provisionally by analyzing data from a subject who developed iliotibial band syndrome, a common overuse injury [14], several weeks after the original study. To track this subject's recovery, we recorded one trial of indoor walking over six days following injury diagnosis (Friday, then Monday through Friday of the following week). In Figure 4, the magnitude feature and raw pattern feature vectors are plotted over four days during the original collection (without induced asymmetries) and over six days during injury recovery. The subject reported the most discomfort and exhibited limping in the first two collections (Friday and Monday). The gait asymmetry features are consistent with the subject's self-reported discomfort.

Prior to the injury, the magnitude feature from unweighted bouts in Figure 4 shows considerable stability over many days. After the injury, however, the magnitude feature becomes negative, indicating a left foot asymmetry. During the course of recovery (around six to seven days following the injury), the magnitude feature returns to its pre-injury baseline. The multidimensional pattern features show a similar result. Immediately after the injury (days +1 and +4), the pattern feature vectors are clearly different from their pre-injury state. Notably, the pattern features show the greatest deviation from normalcy on day +4, when the subject reported the greatest level of discomfort

(a)

(b)
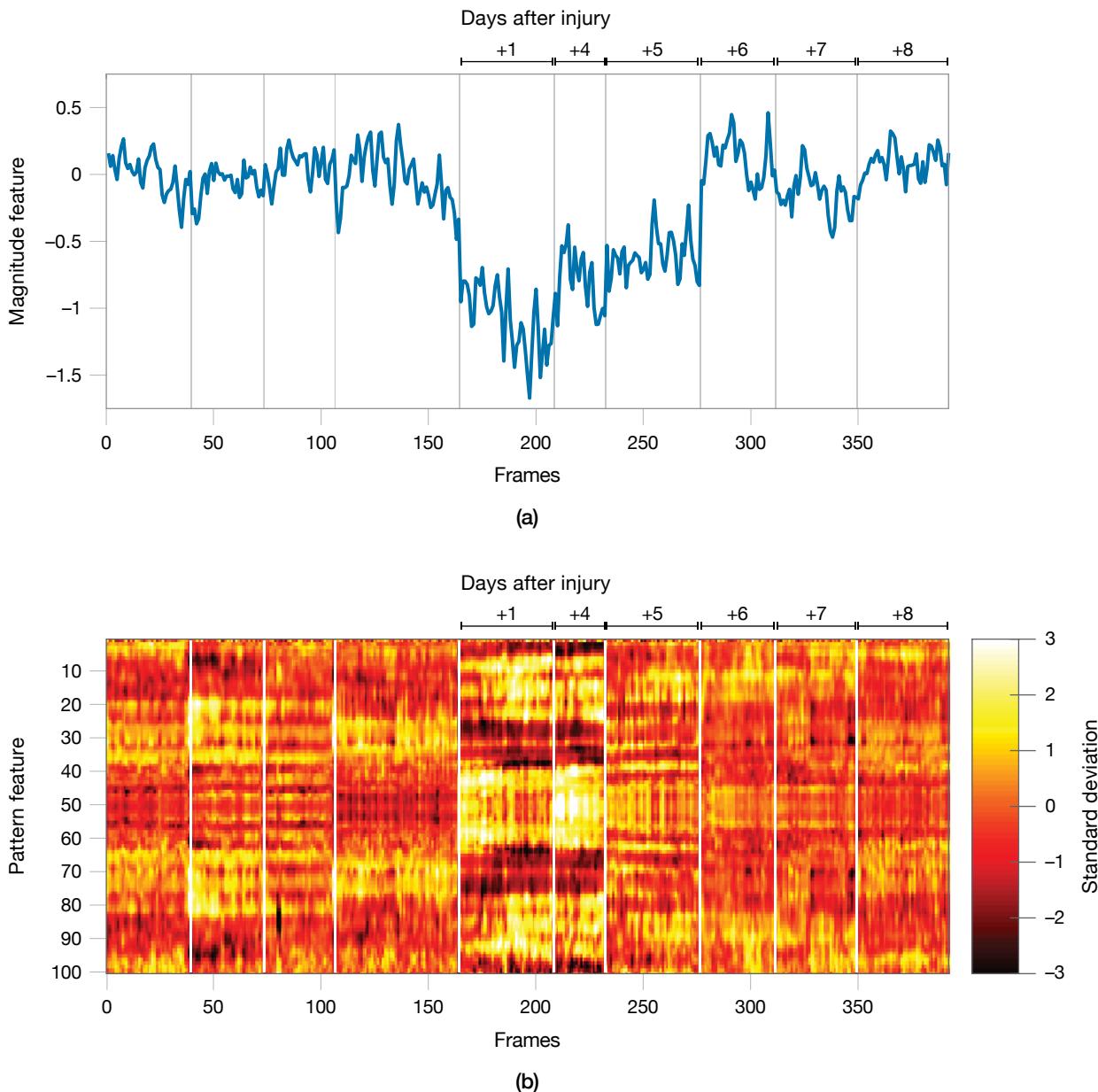
FIGURE 4. Frame-based magnitude feature values for a subject are shown before and after the repetitive stress injury (a). Magnitude features are plotted in units of standard deviation from the mean (following cross-frame normalization). Also plotted are the differences between the raw pattern feature vectors of the left and right feet (b). The detection algorithm uses principal component features, which are extracted from these high-dimensional vectors. The y-axis of Figure 2, for example, shows the first principal component of the pattern features. The changes in magnitude and pattern features correspond to the timing of the subject's injury, indicating that the features tracked in this study have the potential to provide insight into real injuries.

and decided to stop walking halfway through the trial. This small-scale case study suggests that the feature set presented in this article has potential for detecting and tracking actual MSIs and subsequent recoveries.

Estimating Load Carriage from a Commercial Torso-Mounted Accelerometer

Heavy loads increase the risk of MSIs for foot soldiers. Military personnel commonly engage in training and operational activities in which they carry heavy loads (77 to 143 pounds or more) that increase the risk of MSIs [15]. Real-time continuous monitoring of load carriage would improve the ability to assess these risks. In addition, characterizing the effect of load on gait dynamics could allow early detection of MSIs or thermal work strain by analyzing unusual gait patterns.

We used information from a body-worn accelerometer to develop a load-estimating algorithm. The algorithm extracts features that characterize movement dynamics from a commercial torso-mounted accelerometer and uses statistical models to map the features to load estimates. The algorithm obtains rapid estimates of load with robustness to changes in equipment configuration, walking conditions, and walking speeds. On a combined dataset of soldiers and civilians carrying loads that ranged from 0 to 89 pounds, load estimates were obtained with correlation $r = 0.91$ and mean absolute error <10 pounds.

The analysis focused on a five-kilometer march in which soldiers carried different loads corresponding to different equipment configurations. Each soldier wore different equipment on different days. The total march durations varied between 40 and 68 minutes. The soldier data were augmented by civilian data collected from trials at Lincoln Laboratory in which 31 volunteers (19 men and 12 women) between the ages of 18 and 65 wore the Equivital EQ-02 heart-rate monitor/accelerometer during natural walking. For 26 of these subjects, we compiled a single trial of unloaded walking data from multiple indoor and outdoor walking segments. The outdoor walking took place on a looped 0.4-kilometer gravel path that included uneven terrain and an eight-meter elevation change. Five subjects walked only outdoors, with loads of 0, 20, and (for one subject) 41 pounds. A weighted backpack was the load.

Figure 5 shows the raw acceleration data from a single soldier bearing a moderate load of 45.2 pounds (a) and a heavy load of 84.2 pounds (b). The plots represent 10 seconds of a 60-second data frame that has been converted to standard units (z-scoring) with three time series vertically offset for easy viewing.
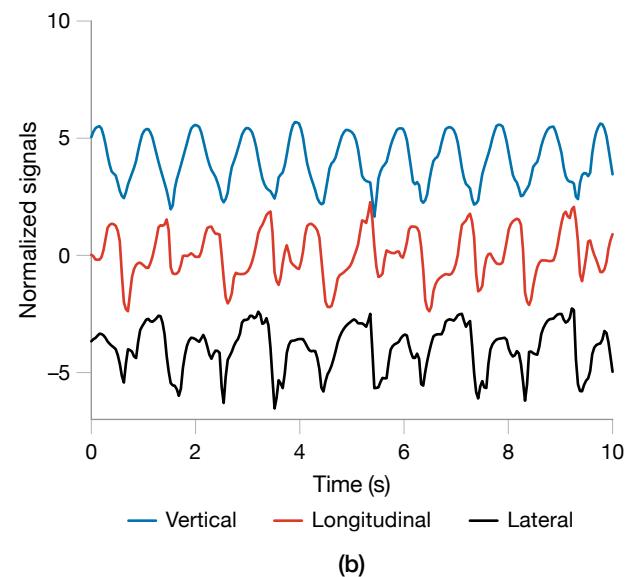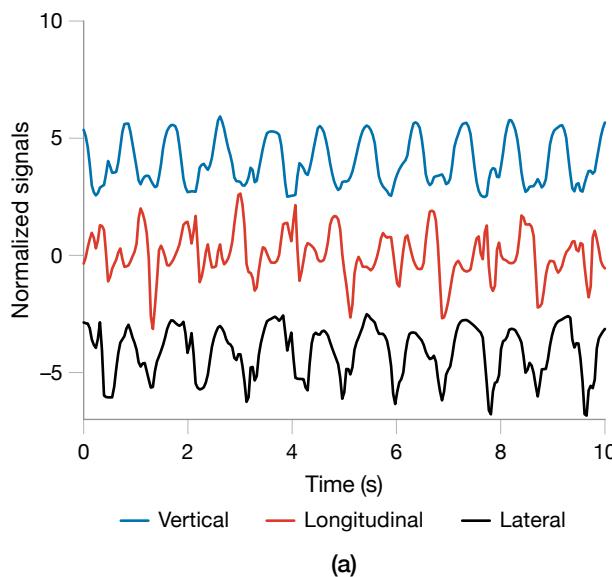
FIGURE 5. The graphs show acceleration signals from the same soldier carrying a load of 45.2 pounds (a) and 84.2 pounds (b). The signals are plotted in standard units (with 0 being the mean), z-scored within 60-second frames, and offset for easy viewing. Vertical, longitudinal, and lateral refer to the axes of measured acceleration relative to the torso. Autocorrelation shape features, shown in Figure 6, can be used to make inferences about the effect of load carriage on acceleration for each axis.
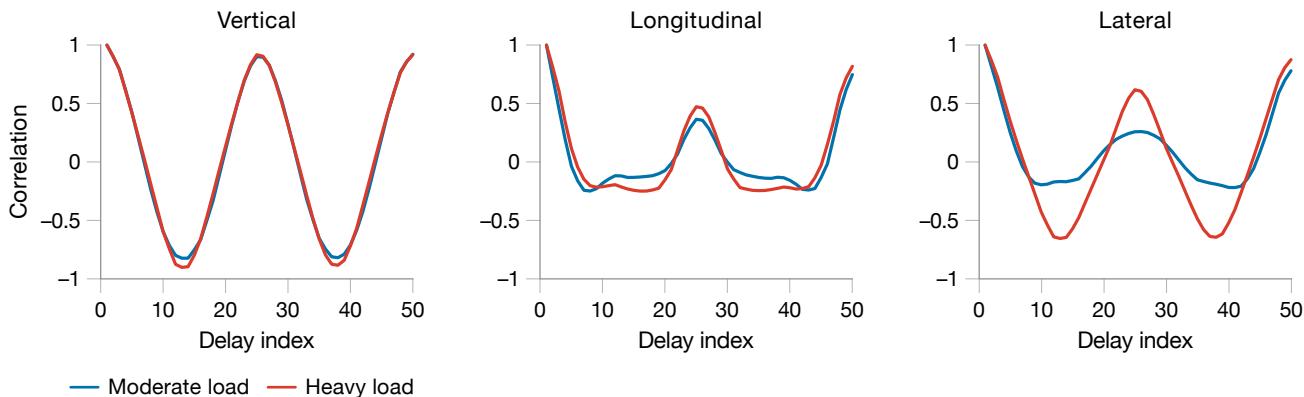
FIGURE 6. Depicted are the autocorrelation shape features, which are the time-scaled autocorrelation patterns for the three axes described in the data frames illustrated in Figure 5. The longitudinal and lateral axes show the largest changes between the moderate (blue) and heavy (red) loads. These results indicate that increasing load carriage causes the largest changes to the dynamics of acceleration in the longitudinal and lateral movement directions.

Autocorrelation shape features that are based on time-scaled autocorrelation patterns are used to analyze the dynamics within each accelerometer axis separately, (Figure 6). These autocorrelation patterns are plotted over a range of time delays that span the average duration of a single stride. The differences between the moderate load (blue) and heavy load (red) autocorrelation patterns, particularly in the longitudinal and lateral axes, indicate a basis for discriminating load levels.

We use correlation structure features to holistically characterize relationships across the three acceleration axes and across time delays. This analysis is done by constructing high-dimensional correlation matrices by employing time-delay embedding, such that each matrix represents a full set of correlations within and across acceleration axes at a set of relative time delays. We then use the matrix eigenspectra to quantify torso dynamics properties. Figure 7 shows that the correlation structure features (the matrix eigenspectra) obtained from a heavy load (red) differ from those obtained from a moderate load (blue) for the same subject. The reduction of lower-rank eigenvalues with load indicates a reduction in torso dynamics with fewer independent modes of motion.

The autocorrelation shape features and the correlation structure features were mapped into load predictions by using a regression algorithm, which is trained on held-out subjects by applying leave-one-subject-out cross-validation [16]. The load-estimation algorithm is applied to a combined dataset consisting of loads ranging between 0 and 89 pounds. The algorithm is able to predict loads with high accuracy, producing estimates with a mean absolute error of 9.57 pounds and a Pearson correlation with true loads of $r = 0.91$. Figure 8 shows a scatter plot of estimated load as a function of true loads, with a linear regression fit to the estimates shown in red. Despite the high overall correlation between estimated and true load, some error in the load estimates remains. Further testing has indicated that much of this remaining error may be attributed to cross-subject differences in body morphology that result in different levels of change in torso dynamics given the same absolute load. Excessive trunk swaying from fatigue or an unbalanced load may also hinder the algorithm's ability to estimate load. A path for future research is to automatically calibrate the load algorithm for each individual, allowing the algorithm to more accurately track within-subject changes in load over time.

The techniques introduced in this section, which characterize complementary properties of accelerometer movement dynamics, could prove useful for other applications, such as early detection of MSIs, detection of excessive thermal work strain, and monitoring of physical fatigue. We extended this work to consider estimating biomechanical state characteristics from a wide variety of sensors.

Multimodal Measurements for Biomechanical Analysis

Body-worn accelerometers can be used to detect gait acceleration differences and can give an estimate of load carried, but more information is needed to truly

understand and interpret more complex and long-duration dynamic movement, namely accelerations and rotations of the feet and ankles, GRFs, and characteristics of the local terrain. While wearable sensors exist for each of these three measurements, no commercial system combines the three modalities for lower-leg biomechanical measurements, and the analytical tools needed to comprehensively integrate the multimodal sensors and interpret lower-leg movement are still an area of open research.

Recently, sensors embedded within or near footwear [17] have increased the possibility for field GRF measurements [18]. Several of these technologies include modified insoles, such as pressure-sensitive fabrics [19], force-sensitive resistors [20], and capacitive arrays [21]. These approaches have excellent packages and are compatible with many types of footwear, but they are limited in dynamic range, sampling frequency, and accuracy. In-shoe technologies often use small load cells and force plates integrated around a sandal [22, 23] or underneath a soldier's boot [24]. While the in-shoe approaches tend to demonstrate higher accuracy and more robustness compared to insole measurements, they are expensive, and their large size and weight tend to alter the normal gait of the user. These commercial and research-based load measurement systems represent progress toward ambulatory sensing; however, they are not complete systems and lack the integration of all the sensing modalities needed for measuring lower-limb biomechanics. To truly understand lower-leg biomechanics, GRF and movement sensors need to be combined into a portable, nonobtrusive format.

Direct measurement of GRFs allows researchers to calculate the dynamic forces of the foot. These dynamic forces, in combination with foot and ankle motion measured by IMUs, may allow the calculation of dynamic forces around the ankle joint. GRF and terrain measurements could also play a role in estimating energy expenditure and inform strategies for load carriage configurations that minimize injuries and maximize performance over a variety of terrains.

To address the gap in multimodal sensors and field measurements, we developed a lower-leg biomechanical measurement system called the Mobility and Biomechanics Insert for Load Evaluation (MoBILE). The MoBILE system combines a shoe insert that captures foot measurements with an ankle sensor. The insert measures
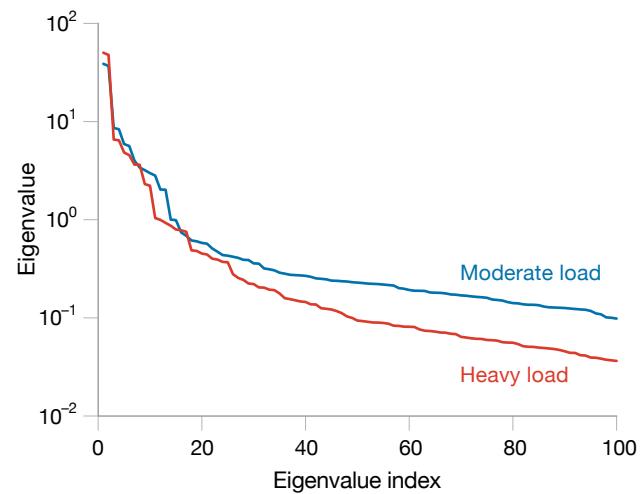
FIGURE 7. Depicted above are the channel-delay correlation matrix eigenvalues for a moderate load (blue) and for a heavy load (red). The heavy load causes a reduction in magnitude at the low-rank eigenvalues.
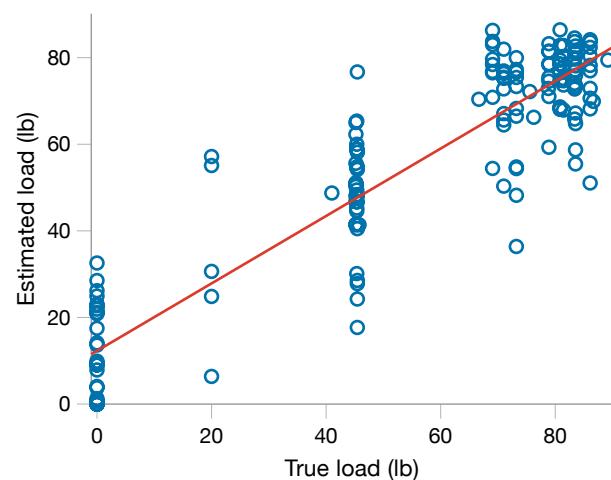
FIGURE 8. The plot shows the estimated load as a function of true load for the combined dataset consisting of 125 soldier trials (≥45 pounds) and 37 Lincoln Laboratory trials (≤42 pounds). The algorithm obtained predictions with a mean absolute error of <10 pounds and correlation >0.9. The plot demonstrates a high overall correlation between estimated and true load, though some error remains because of factors such as variability between subjects' bodies.

GRFs at the toe, arch, and heel of each foot by using custom-designed, inexpensive, flat, and flexible load cells (Figure 9). These novel load-sensing elements have a large dynamic range (up to 3,530 Newtons or 800 pounds) and measure vertical GRFs at high sampling rates. In addition to having GRF sensors, MoBILE also incorporates

nine-axis IMUs (three axes each for the accelerometer, gyrometer, and magnetometer) at each foot and above each ankle and a barometer in the ankle sensor package to measure local terrain features (Figure 9).

The load-sensing elements in each insert, constructed with strain gauges, measure ground resistance in response to changes in load. To convert resistance to load, we used an Instron loading apparatus to determine static (standard curve) and dynamic (creep test) material property parameters for each sensor. The material properties can be used in a viscoelastic model to characterize the viscous and elastic material components to account for the influence of the rubber and foam insert material on the output resistance [25].

The compact MoBILE ankle package integrates the sensors with readout electronics, high-rate data storage, and a rechargeable battery to enable offline analysis of locomotion without altering the mobility of the wearer. We used the system to characterize stationary measurements of weight and dynamic responses (e.g., during walking and running) and directly compared the results to measurements of a force-sensing treadmill. We are also developing analysis tools to evaluate changes in biomechanical gait state during prolonged exercise or training events, and we are developing data-driven algorithms to interpret the large quantity of measured data. Our goal is to characterize an individual's biomechanical state and detect precursors of MSIs in the lower limbs before load-related injuries are incurred.

MoBILE Dataset

The MoBILE system records a rich, high-dimensional set of data. A nominal recording includes data from two sets of load sensors (toe, arch, and heel sensors on each foot), four nine-axis IMUs (one on each foot and each ankle), and two barometric pressure sensors (one on each ankle). This configuration results in 44 independent and concurrent data streams available for analysis. Most of the MoBILE sensors are sampled at rates that are compatible with those of laboratory-grade sensors, with an overall data rate of more than 7,000 samples per second (Table 1). The MoBILE system yields high-resolution data that are recorded and stored for algorithm development, state classification, and anomaly detection.

The data analysis incorporates many well-established gait features from the literature, including cadence and
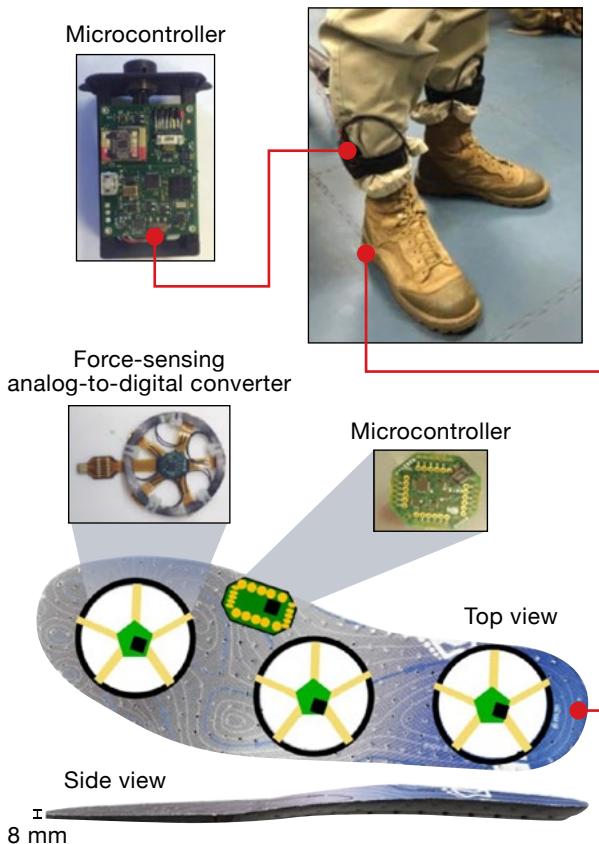
FIGURE 9. Lincoln Laboratory has developed the Mobility and Biomechanics Insert for Load Evaluation (MoBILE), which includes a nine-axis inertial measurement unit for each foot and ankle and a left and right shoe insert that measures ground reaction forces at the toe, arch, and heel of each foot. Each insert and ankle unit contains a microcontroller unit and an analog-to-digital converter and associated electronics.

foot-contact time. The addition of concurrent data from multiple sensors enables the investigation of many new features, such as the peak forces of the heel and toe, and the relative timing of events within the gait period. The rich datasets support the exploration of many more load-based features, ranging from timing and magnitude values within and across the sensors, to symmetry indices similar to those explored in accelerometry, to statistical and information-theoretic features like kurtosis and entropy that characterize the shape or content of the signal waveforms.

Data Preprocessing

After downloading the raw set of measurements from MoBILE, we segmented the data into 10-second windows with five seconds of overlap. Within each window, two

main types of features are extracted: window based and stride based. Window-based features are summary statistics of the entire 10-second duration of a window, such as mean foot acceleration magnitude. These features characterize overall trends in the sensor values without explicitly segmenting the individual strides. To extract intra-stride events, such as the heel strike and toe off, it is necessary to determine which windows contain ambulatory motion. Windows with ambulatory motion are detected by setting threshold levels on the foot gyrometer energy and standard deviation. When the thresholds are exceeded, a window is classified as active, and the start of each individual stride is detected. Stride-based features are calculated for each individual stride in the 10-second window and reported as summary values (e.g., mean, standard deviation, median) within that window. An example of a stride-based feature is mean resistance at the heel strike. The sum of all three load-sensing elements is used to parse the load data into strides, and the foot acceleration magnitude is used to parse the IMU data. Both stride-parsing algorithms use a template-matching scheme to identify individual steps. When truth data, such as the start time, rest periods, trips, and falls, are available, truth annotations are overlaid on the data to support analysis and the development of supervised classifiers to predict anomalous gait patterns.

When biomechanical data are measured over a short baseline period (a few minutes), a normal gait characterization, including nominal gait and weight distribution features, can be established for an individual. Deviations from the baseline may provide insight into the individual's physical status to inform injury avoidance, mission readiness assessment, or performance enhancement. With the goal of characterizing meaningful biomechanical changes over time, we are conducting research to understand which features are important for predicting health and performance or for providing actionable feedback on an individual's physical state.

Activity Maps

As a way to get an overview of the data, activity maps can be generated to visualize trends across the data streams throughout the gait cycle. One way to construct an activity map is to segment the data into distinct events, such as portions of a data collection. Once segmented, the data can be visually inspected for changes or for the emergence of trends when different conditions are compared (for example, walking with and without a heavy load). This comparison can provide guidance on which data streams and/or time points may be strong candidates for feature extraction in the subsequent analysis.

As an example, Figure 10 shows activity maps during a walk with summary data from different time points during the walk. The maps show accelerometer and gyrometer data from the left and right feet and left and right ankles. In this example, the acceleration and gyration standard deviations are changing during the course of the walk.

Deriving Altitude Rate from Barometric Pressure

The barometric pressure sensor in the MoBILE system tracks relative changes in barometric pressure over time. Although in theory, barometric sensor measurements could be converted to absolute elevation changes, some challenges make the conversion impractical. In particular, understanding the relationship between barometric pressure and absolute elevation relies on knowing the local atmospheric pressure, which can change rapidly over time and space. Rather than aiming to estimate absolute elevation, a more practical approach is to estimate relative altitude changes over short intervals

Table 1. Summary of MoBILE Sensors and Data

SENSOR TYPE	AXES	NUMBER	TOTAL	SAMPLING RATE
Load	3	2	6	Up to 600 Hz
Inertial measurement unit	9	4	36	100 Hz
Altimeter	1	2	2	15 Hz

of time during which the local atmospheric pressure is assumed to be constant. Calculating a moving average of relative altitude change or altitude rate provides important context about the terrain.

Figure 11 illustrates the derived altitude rate from the MoBILE system for a round-trip march over a hilly path and compares it to a reference elevation profile extracted from Google Earth for the same hilly path. Notably, the altitude rate calculation gives reliable estimates of when the subject is walking on flat terrain versus walking up or down moderate or steep hills. This information provides valuable context for interpreting lower-leg motion features.

MoBILE Inertial Measurement Unit Data

As previously described, we can detect gait asymmetry from foot-mounted accelerometers [26]. The inclusion of nine-axis IMUs in the MoBILE system allows for the expansion of gait analysis to include foot gyration and ankle acceleration, gyration, and magnetometer data. The concurrent measurement of 36 axes of motion in the MoBILE system leads to the extraction of hundreds of IMU-based features.
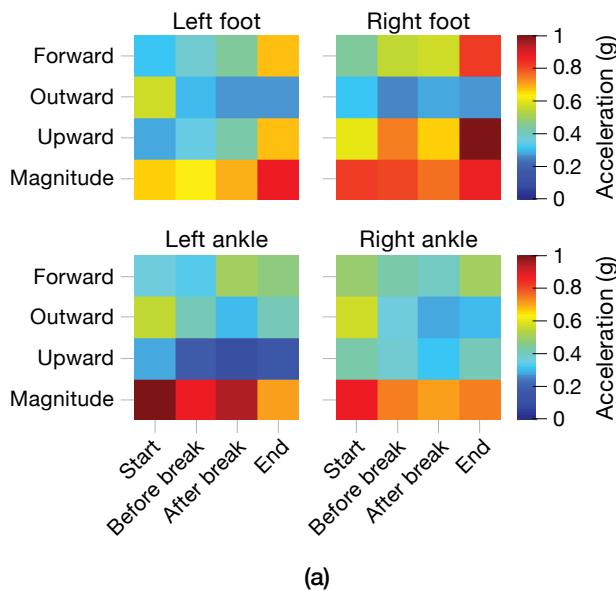
Window-based features such as the mean, median, and standard deviation of each axis (and associated magnitudes) can provide insight into general trends for each sensor over time or in response to relatively slow-changing conditions. Stride-based features allow for tracking more precise changes, such as the value of each sensor at stride landmarks (e.g., heel strike or toe off). Parsing the data into strides also supports the calculation of commonly used timing features, such as foot-to-ground contact time, cadence, percentage of time in stance, and percentage of time in swing. Figure 12 shows an example data window and the average stride across the window for each axis of the foot accelerometer.

After extracting IMU window- and stride-based features, we can perform additional analyses and comparisons. Symmetry, or lack thereof, between right and left gait features is a powerful indicator of gait anomalies (e.g., tripping) or gait asymmetries. The emergence of a gait asymmetry over time may be indicative of changes in stability, mobility, agility, or injury state. To further explore the indications of gait symmetry, we calculated symmetry indices as defined by Equation (1) for each IMU feature.
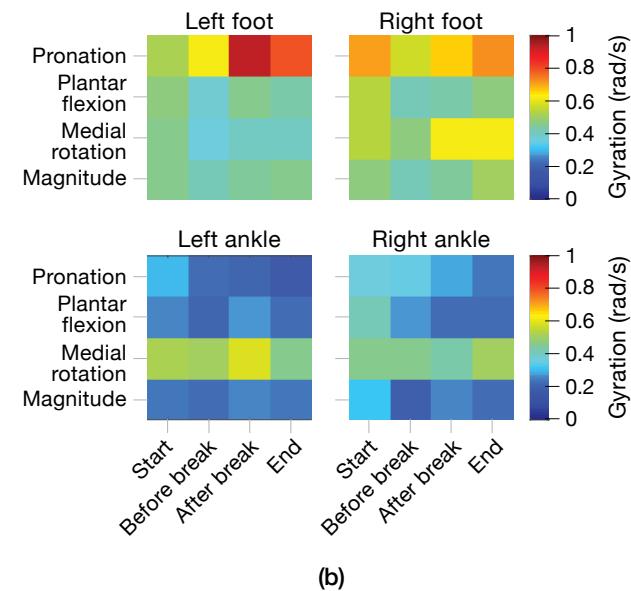
FIGURE 10. Example foot IMU sensor activity maps are depicted. The plots in (a) show the acceleration axis at different points during a march. The plots in (b) show the gyration axis at different times during a march. Pronation, plantar flexion, and medial rotation refer to different states in which the foot is flexed or rotated. The acceleration patterns show a difference in magnitude between the two feet throughout the data collection. The gyration pattern indicates a sharp change in the left foot pronation (collapsing of the foot arch upon contact with the ground) during the second portion of the walk.
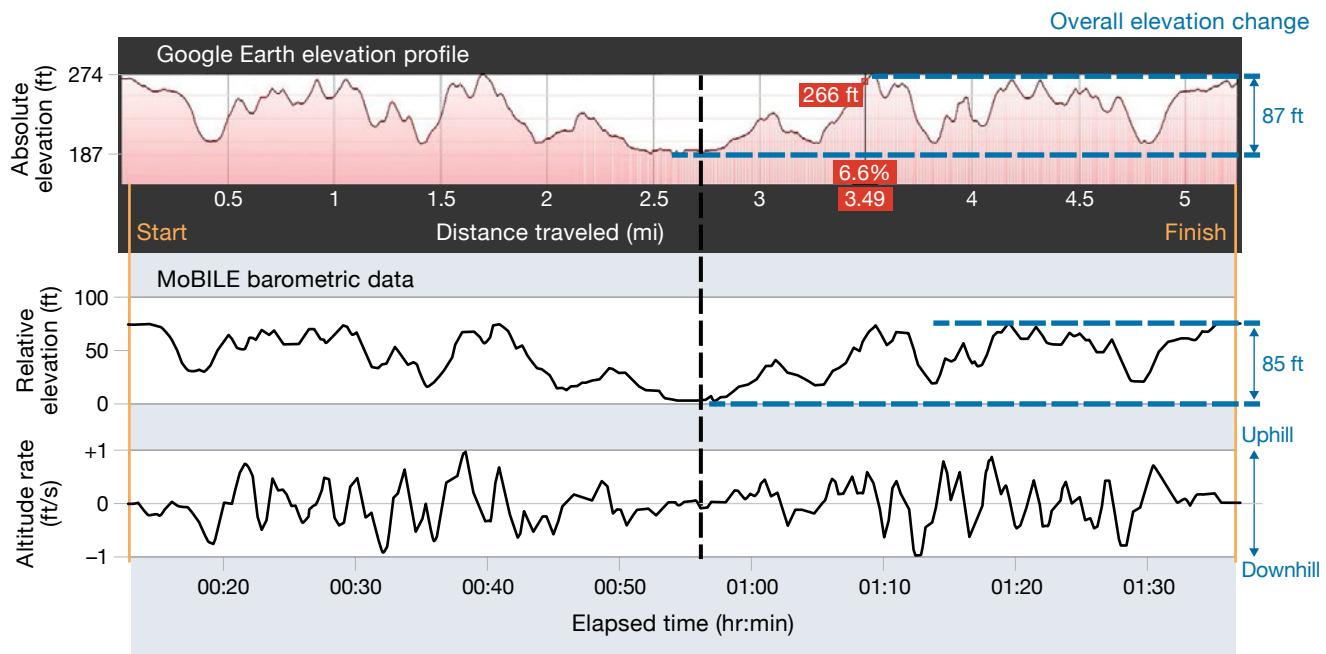
FIGURE 11. The figure shows the comparison of the MoBILE pressure sensor and altitude rate calculations with the Google Earth elevation profile. The MoBILE barometric data match the Google Earth elevation profile, indicating that the MoBILE sensor can provide a reliable picture of the terrain traversed by the user.
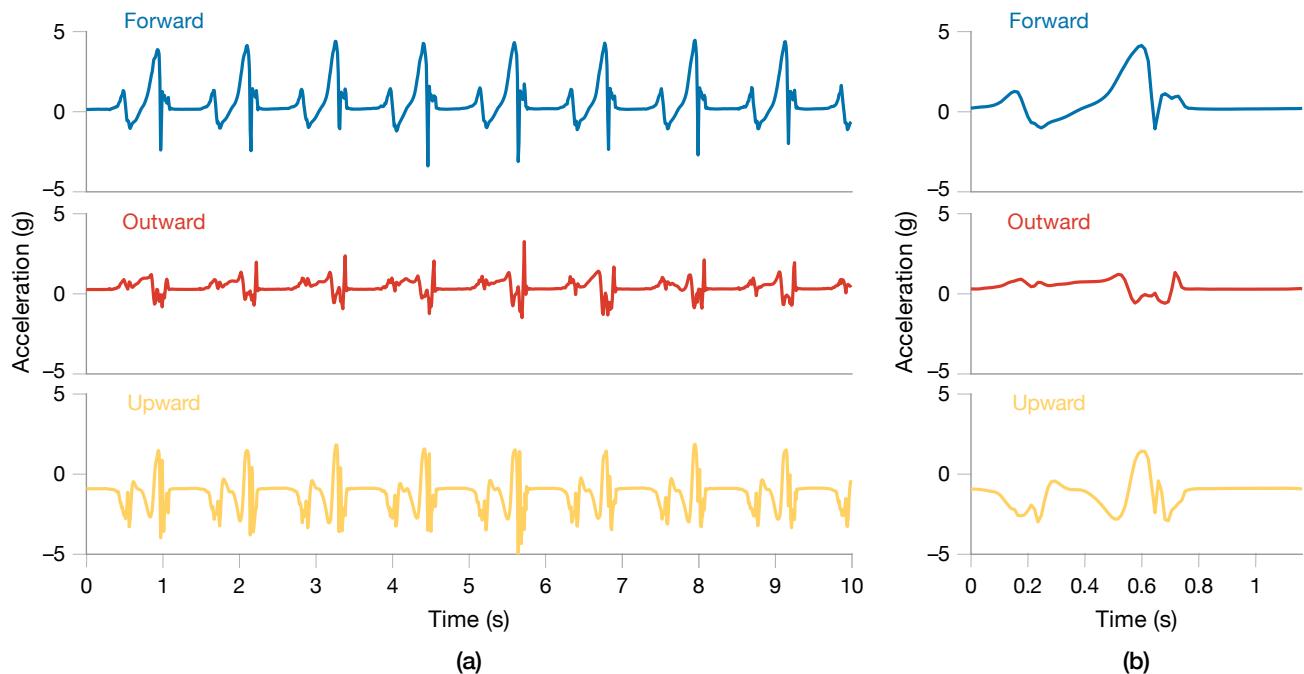
FIGURE 12. The graphs show an example of raw foot-accelerometer data gathered from the three axes of the accelerometer. The traces in (a) represent typical data seen in a 10-second window, and the traces in (b) show the accelerations for one average stride.

The IMU features can also be plotted in relation to condition changes, such as the change in altitude over time (altitude rate). Comparing features to altitude rate highlights terrain dependencies. Identifying such correlations can help deconvolve confounding factors, such as changes in gait due to terrain features, in later analysis involving classifiers. In Figure 13, the foot contact time (FCT) and plantar flexion gyration, both stride-based features derived from the foot IMUs, are plotted as a function of altitude rate. Each point on the plot represents a 10-second window during a 10-kilometer march across hilly terrain. The mean FCT has a strong increasing trend versus altitude rate, while the plantar flexion shows a strong decreasing trend versus altitude rate.

Comparing the time evolution of IMU features' baseline values for an individual has the potential to reveal patterns in the data that correlate with biomechanical state changes associated with conditions such as heavy load, fatigue, and varying terrain. Figure 14 shows four example foot IMU features versus load. To avoid the confounding impact of terrain change, the feature comparisons are restricted to windows in which the altitude rate is very small, consistent with the flat region in Figure 13. The symmetry indices show little to no change between a baseline walk (collected the day before the march with no external load) and the initial segment of a 10-kilometer march with a 40-pound load. The symmetry indices do change, however, after the subject walks for an extended period (approximately one hour) with a load. Other features show a significant change from a baseline walk to a walk with load while also showing further change after prolonged walking with a load. All of these features may indicate physical fatigue.

Future IMU work will include additional feature extraction, identification of the best candidate features, collection of more and larger datasets, development of classification algorithms, and multisensor data fusion. In addition, we will incorporate the ankle IMU data and the three magnetometer axes from each IMU into the analysis. Adding the ankle IMU data will allow for the calculation of ankle-specific metrics, inference into the status of the ankle joint, and correlations between the foot and ankle IMU measurements.

(a)

(b)
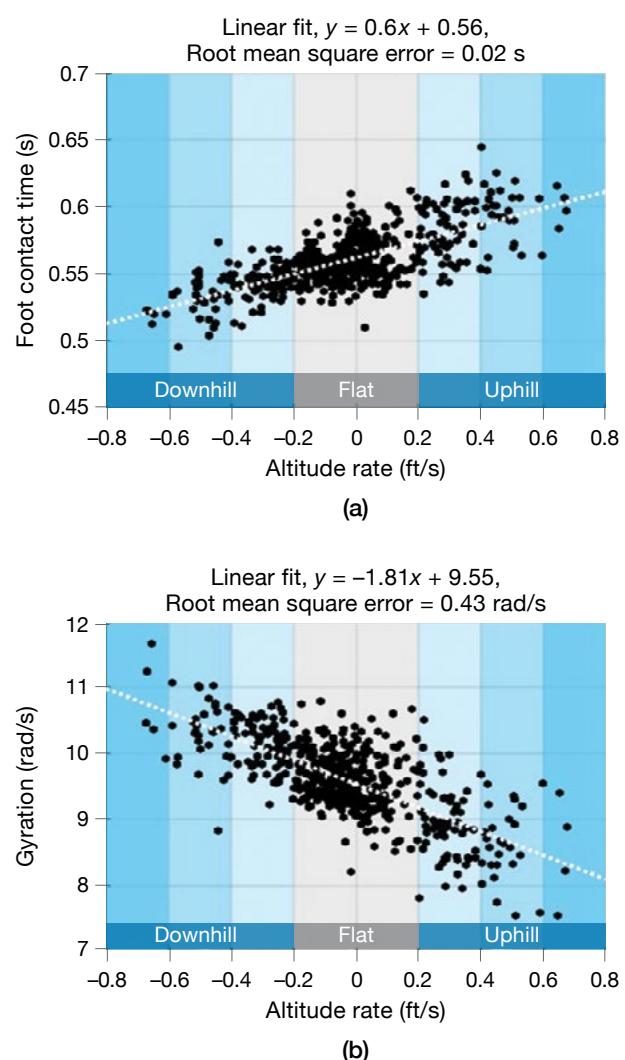
FIGURE 13. The plots show IMU features versus altitude rate and the associated root mean square error (RMSE). The foot contact time (FCT) and gyration vary depending on the terrain feature. For example, the FCT increases and gyration decreases during an uphill walk. The FCT decreases and the gyration increases during a downhill walk. These results indicate that the user is moving slower and has less foot rotation while walking uphill and is moving faster with greater foot rotation while walking downhill.

Load Sensor Data

The raw data from the load-sensing elements, prior to calibration, measure the strain gauge resistance in ohms. The load sensors at the heel, arch, and toe in the MoBILE insert capture the dynamics of the weight distribution across each foot. Figure 15 shows an example of the resistance measured at each sensor during several steps. Six
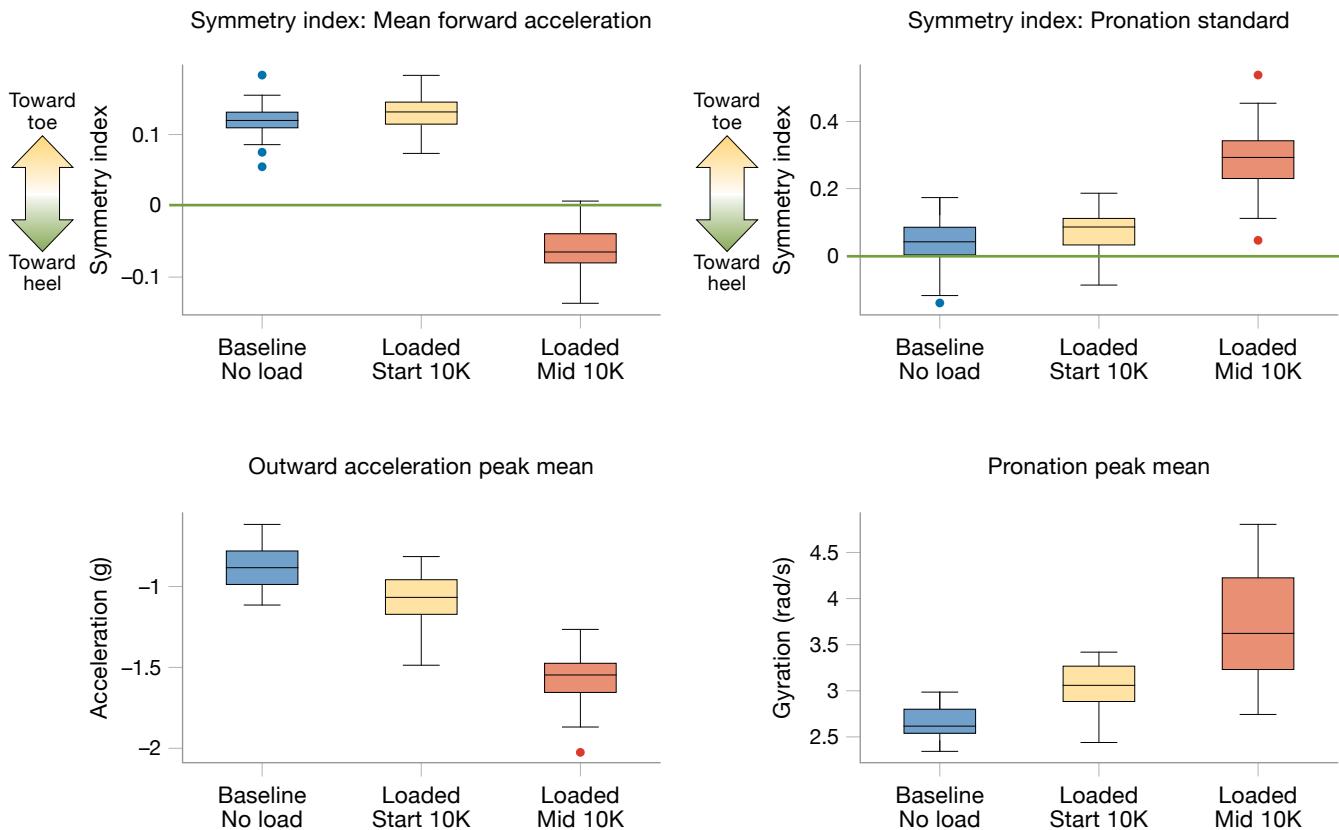
FIGURE 14. Depicted are foot IMU features versus load for a 50-step walking segment (N = 50). The symmetry indices for mean forward acceleration and pronation standard deviation show little to no change during the no-load baseline and the start of a walk with a heavy load. There is a shift in both features after the subject has walked for approximately one hour. Other features, such as outward acceleration and peak mean pronation, show a steady change between no-load baselines, the start of the loaded walk, and after walking with a load for an hour.

consecutive gait cycles from a 10-second frame of data have been overlaid, showing only minor variations of weight distribution from one stride to the next within a frame. The heel, shown in red, is the first point of contact in the step and forms a sharp peak, referred to as the heel strike. Approaching mid-stance, the weight shifts from the heel to a balance between the toe and heel. Although the arch sensor is active throughout most of the stride, it bears only a small portion of the weight compared to the toe and heel. At the end of the stride, the toe sensor forms a peak as the foot begins to unload in an event known as the toe off. Conventional force plates measure a load profile similar to the black lines, which represent the sum of all three sensors.

We can infer many important gait characteristics, including the stride timing and the peak force associated with the heel strike and toe off, from the overall weight profile. The addition of sensors in different regions of the shoe insert also provides a unique opportunity to detect anomalies in weight carriage in the heel, arch, or toe as physical and physiological conditions change. We estimated the resulting force from each sensor by using a viscoelastic model and the time-dependent deformation (creep) properties found during calibration of the insert [30].

Similar to features of the IMU, features derived from the load sensor insert were strongly correlated with changes in altitude rate (uphill, downhill, flat). Figure 16 shows the peak heel and toe force (after calibration) from a 10-kilometer road march in which a 188-pound subject wore a 40-pound vest. The heel strike data show a strong decreasing trend versus altitude rate. The toe sensor is
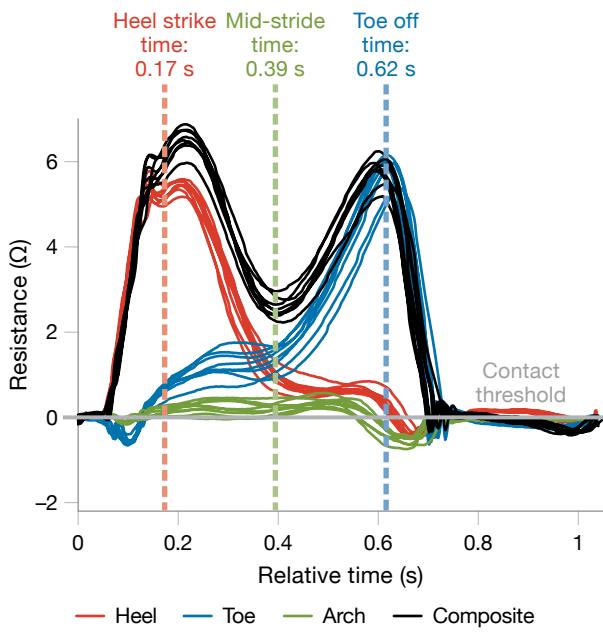
FIGURE 15. The figure shows the MoBILE load sensor resistance measurements from one foot. Data from six gait cycles in a 10-second frame are shown. At the beginning of the stride, resistance peaks at the heel, then shifts to a balance between the heel and the toe, and finally forms a peak at the toe at the end of the stride. The arch bears the smallest portion of the weight. The lines represent the sum of all three sensors.

FIGURE 16. The plots show load distribution as a function of elevation. The heel (a) and toe (b) force data from the load sensor insert were strongly correlated with changes in altitude rate. These results indicate that the subject's weight shifted while walking uphill, downhill, or on flat terrain. The measurements of weight distribution on the feet as the terrain changes may be a good indication of the type of load carriage configuration or fatigue. Models will need to interpret and account for weight distribution as it relates to altitude rate to distinguish changes in weight distribution due to terrain from changes due to fatigue.

less sensitive to terrain, and its data show only a slight increasing trend with altitude rate. Tracking weight distribution on the foot as elevation changes may be a good indicator of load carriage configuration or fatigue. In addition, any analysis will need to understand and account for weight distribution as it relates to altitude rate to separate terrain-related changes from changes that result from fatigue or injury.

Each insert measures load at three points under the foot. Using data collected by these inserts, we can estimate and monitor a center of pressure calculation throughout a collection. The center of pressure is a function of the load at different points on the foot in relation to a reference line (center of insert) as illustrated in Figure 17.

Center of pressure (CoP) is calculated at time t as

$$CoP(t) = \frac{d_{\text{toe}}\, f_{\text{toe}}(t) + d_{\text{arch}}\, f_{\text{arch}}(t) + d_{\text{heel}}\, f_{\text{heel}}(t)}{f_{\text{toe}}(t) + f_{\text{arch}}(t) + f_{\text{heel}}(t)}$$

where d_x is the distance from the center line of the insert

to the center of sensor x, and $f_x(t)$ is the force measured at sensor x at time t. The variable x here indicates the toe, arch, or heel sensor.

Following the stride-based feature approach used in the IMU, we parsed the load data into individual strides and estimated the center of pressure for each stride on the basis of the heel-strike force versus the mid-stride arch force and the toe-off force. Examples of the center of pressure mean and standard deviation for each stride window are shown in Figure 18. The left plot shows the significant variation in center of pressure for a subject walking on hilly terrain. This variation is due to the
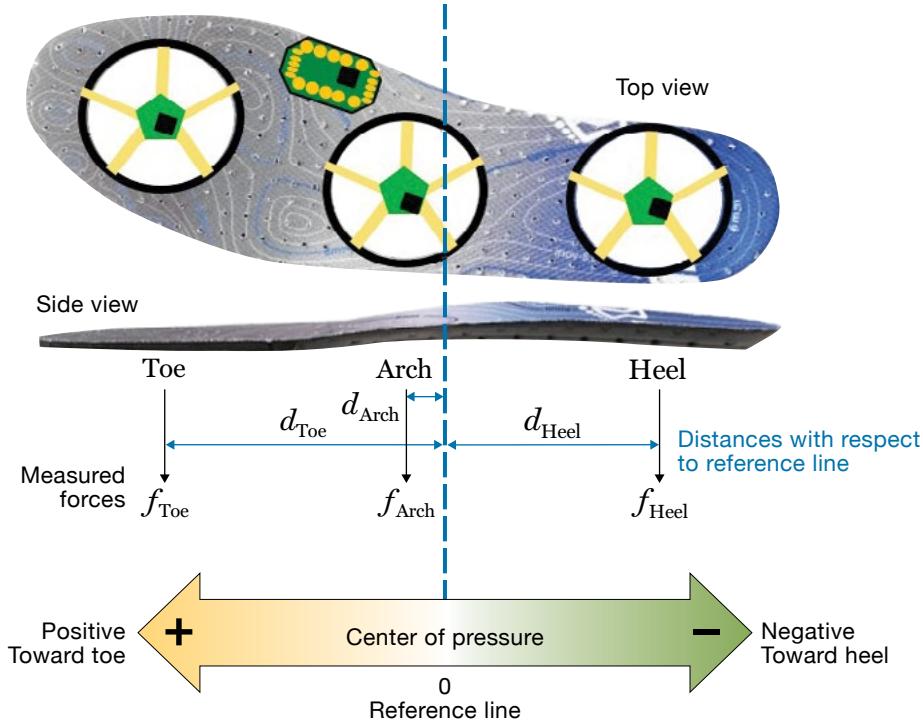
FIGURE 17. The center of pressure calculation for MoBILE is illustrated. When the center of pressure is negative, the user is applying more weight to the heel. When the center of pressure becomes positive, the user is applying more weight to the toes.

redistribution of weight associated with walking up and down slopes. The plot on the right compares three different instances of flat walking for the same subject. A notable shift in the center of pressure occurs as weight is added to the subject, and an additional shift happens after the subject has walked for 30 minutes with the load. While this example highlights elevation and load changes, we could also use this feature to estimate shifts in weight distribution caused by other relevant biomechanical conditions, such as a rebalancing of the load or an injury.

Energy Expenditure

Energy expenditure is measured directly by monitoring oxygen consumption. However, this type of direct measurement can be cumbersome to collect in the field because the equipment to measure energy expenditure is bulky and not user friendly. Alternatively, a number of equations have been proposed to predict energy expenditure by using indirect measurements of parameters, such as speed, body weight, external load, and information about terrain slope and type [27]. We can use measurements from the MoBILE sensor as dynamic inputs to these predictive equations. For example, the Pandolf model with a correction factor for negative slopes [28, 29] and the more recently developed Weyand model [30] rely only on speed, terrain slope, terrain factor, and weight. While MoBILE does not measure absolute speed, the cadence derived from the IMUs, paired with an assumed stride length (height dependent) for an individual, may be a reasonable surrogate for speed in these equations for calculating energy expenditure. Similarly, grade can be estimated from altitude rate and the assumed stride length, and total weight can be measured from intervals in the load sensor data that depict when the subject is standing.

Figure 19 shows estimates of energy expenditure (reported in volume of oxygen consumed per minute per kilogram) for the first hour of the 10-kilometer march with a load. The low energy expenditure at the beginning corresponds to a period during which the subject is standing prior to the beginning of the march. Throughout the march, the estimates of energy expenditure fluctuate, with energy expenditure increasing on the uphill stretches. In the near future, real-time calculation of energy expenditure could be possible with MoBILE or another sensor. Wearable sensors that monitor energy expenditure in real time could be important tools for determining energy intake needs and optimizing physical performance.

Identifying Biomechanical Anomalies from Multimodal Measurements

Multimodal sensors, such as the ones integrated into MoBILE, are capable of measuring subtle biomechanical changes that occur in response to different physical activities. The presence of these measurable changes suggests promising potential for training machine learning algorithms to use the combined features from a rich multisensory dataset to classify the physical state of an individual.

In addition to collecting data with the MoBILE sensor, we also collected heart rate, core temperature, and skin temperature measurements during the 10-kilometer march. Combining load sensors with physiological monitoring sensors could increase the classification potential for a multimodal system. Some potential classes for machine learning might include the weight-bearing status (wearing a heavy pack versus no load), difficulty of the terrain (uphill versus downhill), index of fatigue, efficiency of load carriage, and early indications of lower-limb MSIs. While some changes in the feature values can be seen by eye, machine learning algorithms can learn to recognize subtle changes across many dimensions to provide robust estimates of the physical state.

The radar plot in Figure 20 demonstrates some of the measured differences in an eight-dimensional space that includes features derived from the load sensors, IMUs, barometers, and physiological status monitoring. The
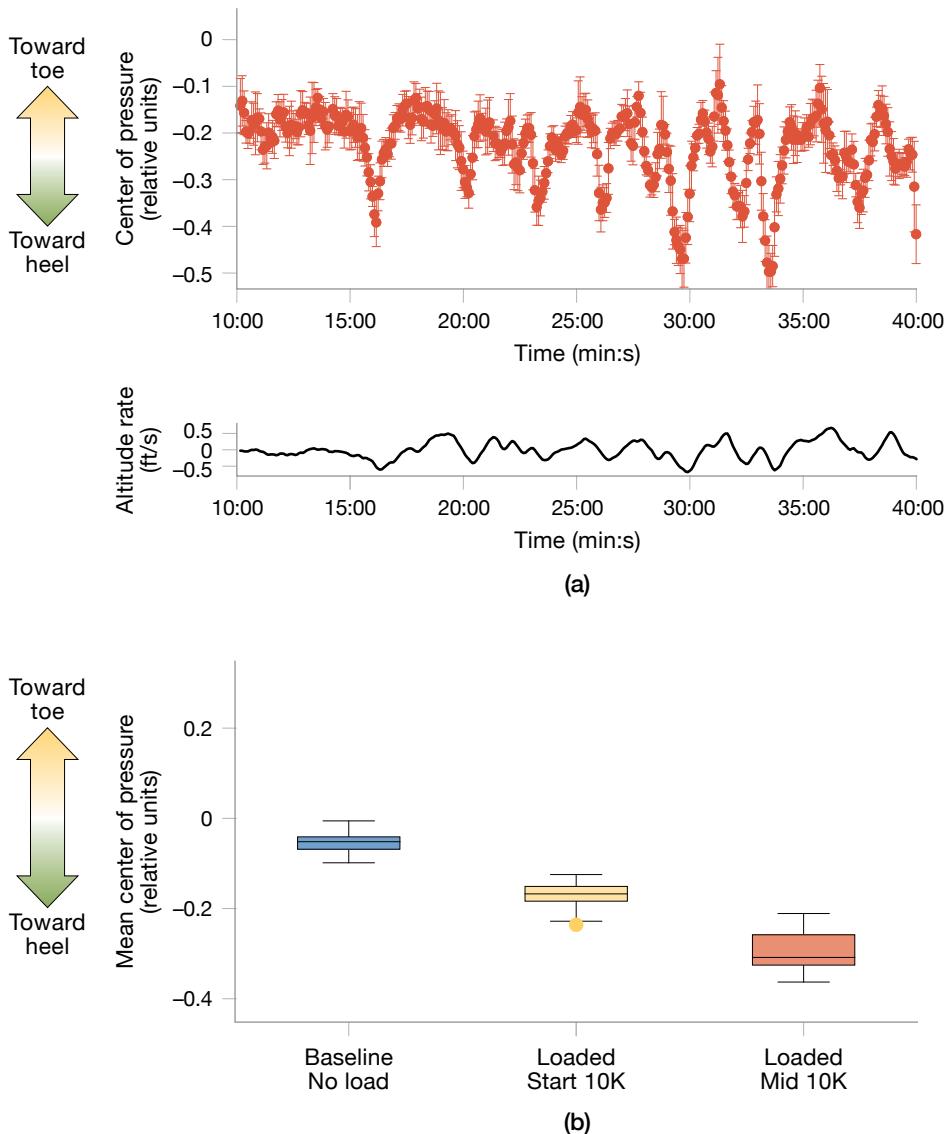
FIGURE 18. The plots show the center of pressure of a subject during a data collection. The two plots in (a) compare changes in the center of pressure to changes in altitude rate. The center of pressure shifts to the toes when the altitude rate increases (uphill walking) and shifts to the heels when the altitude rate decreases (downhill walking). The plot in (b) compares the center of pressure on flat terrain during the course of a 10-kilometer march. The center of pressure changed during the march, resulting in the user carrying more of the weight on his heels. While these methods examine the effects of elevation and load carriage on weight distribution, they could also be used to detect weight distribution caused by physical conditions such as an injury.
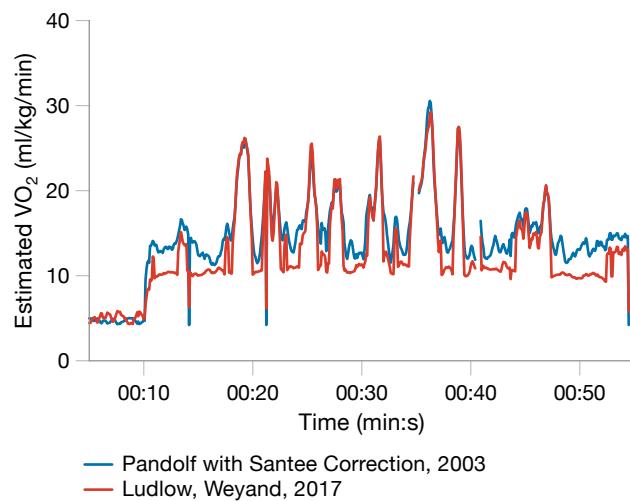
FIGURE 19. The chart, based on MoBILE sensor measurements, shows the dynamic estimate of energy expediture over a 10-kilometer march on hilly terrain. Both the Pandolf and Weyand calculations match closely during periods of high exertion (high VO$_2$) but differ during lower exertion. The difference is due to the fact that the Pandolf equation does not isolate the static part of the metabolic rate when the subject is just standing with the load. This characteristic of the Pandolf equation results in higher baseline metabolic rates.

three colored polygons indicate three different conditions: the baseline (no load), start of the march with a 40-pound load, and an hour into the loaded march, which we assume to correspond to a mildly fatigued state for the individual. Several features, such as the center of pressure, toe and heel force, and forward acceleration, change dramatically over the three conditions. Other features, such as medial rotation, heat strain index, and core temperature, change slightly. The significance of the amount of change for each feature has not been accounted for in this plot. For example, a slight change in core temperature can be just as significant as a large change in the center of pressure. Future work in this area includes selecting a subset of powerful features and building a classifier to predict characteristics of the individual's physical state over time.

Future Work

Going forward, an overarching goal is to translate data-derived features into actionable information that allows the user to assess the information and respond. For example, if a person's gait is significantly altered to accommodate extra load, that information should be reported in an interpretable manner that highlights the
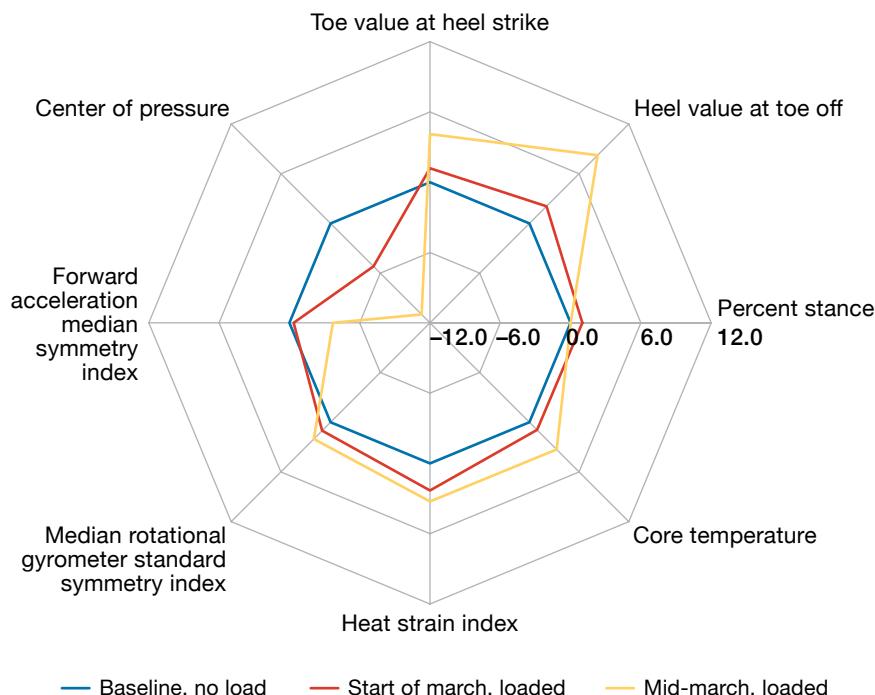
FIGURE 20. The radar plot depicts measured differences from eight features that are derived from the load sensors, IMUs, barometers, and physiological status monitoring. The blue, yellow, and red lines indicate feature values for the baseline (no load), mid-march, and start of march, respectively. Several of the features show little change (percent stance, heat strain index), while others change from the no-load baseline to the start of the march (loaded) and then from the start of the march to mid-march. Future work includes selecting fatigue- and injury-indicating features and building a classifier to predict characteristics of the individual's physical state over time.

risk of injury and helps determine appropriate work-rest cycles or weight redistribution to prevent injury and optimize performance.

The ability to relate biomechanical measurements and models from within a laboratory setting to measurements in real-world environments requires capturing movement data in the field. Generating field datasets from accelerometers, instrumented inserts, or any other sensor or measurement is the first step to understanding and interpreting movement. These data need to be compared to ground truth and over a long duration. Once these data are generated, a wealth of information can be extracted from them. Current commercial systems focus on providing the user with a few snippets of information, but to be truly useful, the measurements and algorithms need to be merged. MoBILE is an attempt to provide accurate biomechanical measurement and to develop the analysis tools to extract features for the classification of movement types and potentially injured states. ∎

References

1. K.G. Hauret, B.H. Jones, S.H. Bullock, M. Canham-Chervak, and S. Canada, "Musculoskeletal Injuries Description of an Under-Recognized Injury Problem Among Military Personnel," *American Journal of Preventatitve Medicine*, vol. 38, no. 1, 2010, pp. S61–S70.

2. A. Yassi, "Repetitive Strain Injuries," *The Lancet*, vol. 349, no. 9066, 1997, pp. 943–947.

3. C.E. Milner, R. Ferber, C.D. Pollard, J. Hamill, and I.S. Davis, "Biomechanical Factors Associated with Tibial Stress Fracture in Female Runners," *Medicine and Science in Sports and Exercise*, vol. 38, no. 2, 2006, pp. 323–328.

4. R.B. Davis, "Clinical Gait Analysis," *IEEE Engineering in Medicine and Biology Magazine*, vol. 7, no. 3, 1988, pp. 35–40.

5. K.K. Patterson, W.H. Gage, D. Brooks, S.E. Black, and W.E. McIlroy, "Evaluation of Gait Symmetry after Stroke: A Comparison of Current Methods and Recommendations for Standardization," *Gait Posture*, vol. 31, no. 2, 2010, pp. 241–246.

6. A. Salarian, H. Russmann, F.J. Vingerhoets, C. Dehollain, Y. Blanc, P.R. Burkhard, and K. Aminian, "Gait Assessment in Parkinson's Disease: Toward an Ambulatory System for Long-Term Monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 8, 2004, pp. 1434–1443.

7. C.B. Redd and S.J.M. Bamberg, "A Wireless Sensory Feedback Device for Real-Time Gait Feedback and Training," *IEEE/American Society of Mechanical Engineers (ASME) Transactions on Mechatronics*, vol. 17, no. 3, 2012, pp. 425–433.

8. K. Aminian, K. Rezakhanlou, E. De Andres, C. Fritsch, P.F. Leyvraz, and P. Robert, "Temporal Feature Estimation during Walking Using Miniature Accelerometers: An Analysis of Gait Improvement after Hip Arthroplasty," *Medical and Biological Engineering and Computing*, vol. 37, no. 6, 1999, pp. 686–691.

9. C. Hodt-Billington, J.L. Helbostad, W. Vervaat, T. Rognsvåg, and R. Moe-Milssen, "Criteria of Gait Asymmetry in Patients with Hip Osteoarthritis," *Physiotherapy Theory and Practice*, vol. 28, no. 2, 2011, pp. 134–141.

10. A. Tura, M. Raggi, L. Rocchi, A.G. Cutti, and L. Chiari, "Gait Symmetry and Regularity in Transfemoral Amputees Assessed by Trunk Accelerations," *Journal of NeuroEngineering and Rehabilitation*, vol. 7, no. 4, 2010, p. 4.

11. H. Sadeghi, P. Allard, F. Prince, and H. Labelle, "Symmetry and Limb Dominance in Able-Bodied Gait: A Review," *Gait Posture*, vol. 12, no. 1, 2000, pp. 34–45.

12. R.O. Robinson, W. Herzog, and B.M. Nigg, "Use of Force Platform Variables to Quantify the Effects of Chiropractic Manipulation on Gait Symmetry," *Journal of Manipulative and Physiological Therapeutics*, vol. 10, no. 4, 1987, pp. 172–176.

13. R. Moe-Nilssen and J.L. Helbostad, "Estimation of Gait Cycle Characteristics by Trunk Accelerometry," *Journal of Biomechanics*, vol. 37, no. 1, 2004, pp. 121–126.

14. F.A. Barber and A.N. Sutker, "Iliotibial Band Syndrome," *Sports Medicine*, vol. 14, no. 2, 1992, pp. 144–148.

15. C.M. Clements, D. Moody, A.W. Potter, J. F. Seay, R.E. Fellin, and M.J. Buller, "Loaded and Unloaded Foot Movement Differentiation Using Chest Mounted Accelerometer Signatures," paper presented at the 2013 IEEE International Conference on Body Sensor Networks, Cambridge, 6–9 May 2013.

16. J.R. Williamson, A. Dumas, G. Ciccarelli, A.R. Hess, B.A. Telfer, and M.J. Buller, "Estimating Load Carriage from a Body-Worn Accelerometer," paper in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2015.

17. C. Liedtke, S.A. Fokkenrood, J.T. Menger, H. van der Kooij, and P.H. Veltink, "Evaluation of Instrumented Shoes for Ambulatory Assessment of Ground Reaction Forces," *Gait and Posture*, vol. 26, no. 1, 2007, pp. 39–47.

18. W. Tao, T. Liu, R. Zheng, and H. Feng, "Gait Analysis Using Wearable Sensors," *Sensors (Basel, Switzerland)*, vol. 12, no. 2, 2012, pp. 2255–2283.

19. A.H.A Razak, A. Zayegh, R.K. Begg, and Y. Wahab, "Foot Plantar Pressure Measurement System: A Review," *Sensors (Basel, Switzerland)*, vol. 12, no. 7, 2012, pp. 9884–9912.

20. S.J. Bamberg, A.Y. Benbasat, D.M. Scarborough, D.E. Krebs, and J.A. Paradiso, "Gait Analysis Using a Shoe-Integrated Wireless Sensor System," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 4, 2008, pp. 413–423.

21. A.B. Putti, G.P. Arnold, L. Cochrane, and R.J. Abboud, "The Pedar® In-Shoe System: Repeatability and Normal Pressure Values," *Gait and Posture*, vol. 35, no. 3, 2006, pp. 401–405.

22. P.H. Veltink, C. Liedtke, E. Droog, and H. van der Kooij, "Ambulatory Measurement of Ground Reaction Forces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 3, 2005, pp. 423–427.

23. H.M. Schepers, H.F.J.M. Koopman, and P.H. Veltink, "Ambulatory Assessment of Ankle and Foot Dynamics," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 5, 2007, pp. 895–902.

24. R.F. Lind, L.J. Love, J.C. Rowe, and F.G. Pin, "Multi-axis Foot Reaction Force/Torque Sensor for Biomedical Applications," paper in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2019.

25. J.J. Skrzypek and W.W. Ganczarski, eds., *Mechanics of Anisotropic Materials*. Switzerland: Springer International Publishing, 2015.

26. J.R. Williamson, A. Dumas, A.R. Hess, T. Patel, B.A. Telfer, and M.J. Buller, "Detecting and Tracking Gait Asymmetries with Wearable Accelerometers," paper in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2015.

27. A.W. Potter, W.R. Santee, C.M. Clements, K.A. Brooks, and R.W. Hoyt, "Comparative Analysis of Metabolic Cost Equations: A Review," *Journal of Sport and Human Performance*, vol. 1, no. 3, 2013.

28. K.B. Pandolf, B. Givoni, and R.F. Goldman, "Predicting Energy Expenditure with Loads While Standing or Walking Very Slowly," *Journal of Applied Physiology: Respiratory, Environmental, and Exercise Physiology*, vol. 43, no. 4, 1977, pp. 577–581.

29. W.R. Santee, L.A. Blanchard, K.L. Speckman, J.A. Gonzalez, and R.F. Wallace, "Load Carriage Model Development and Testing with Field Data," U.S. Army Research Institute of Environmental Medicine Technical Note, 2003.

30. L.W. Ludlow and P.G. Weyand, "Walking Economy is Predictably Determined by Speed, Grade, and Gravitational Load," *Journal of Applied Physiology*, vol. 123, no. 5, 2017, pp. 1288–1302.

Appendix

The Sensorimotor Technology Realization in Immersive Virtual Environments Center

The MIT Lincoln Laboratory Sensorimotor Technology Realization in Immersive Virtual Environments (STRIVE) Center was established in July 2016 to promote collaboration among the government, academic, and medical communities on key challenges in clinical rehabilitation, wearable technology development, and advanced operational training. The STRIVE Center's flagship capability is the Computer-Assisted Rehabilitation Environment (CAREN), one of three such systems in the world that features a 24-foot dome. The CAREN enables the assessment of individuals' cognitive and physiological performance as they interact with a fully immersive virtual environment. It features a 360-degree visualization screen extending from the dome onto the treadmill surface, surround sound, an 18-camera motion-capture system, a dual-belt instrumented treadmill mounted on a six-degrees-of-freedom motion platform, and integrated sensing of electrophysiological signals and vital signs. The system is run by real-time software that allows researchers to introduce elements into the scene in a flexible manner. The facility, approximately 4,000 square feet in size, is operated by the Lincoln Laboratory Human Health and Performance Systems Group as an asset available for use by researchers throughout the Laboratory to support programs spanning all major mission areas.

Research Focuses
Clinical Research

The CAREN system offers a unique platform for clinical research by combining exceptional flexibility and precision of prescribed sensory perturbations with the simultaneous collection of multimodal physiological data. This real-time

The Sensorimotor Technology Realization in Immersive Virtual Environments (STRIVE) Center features a virtual reality dome for research and development across many mission areas.

system allows clinicians and researchers to tailor training and rehabilitation protocols to individual patients. Key focus areas within the STRIVE Center's foundational and applied clinical research programs include balance, movement, and cognitive performance.

Technology Development

Field testing is essential to rigorously evaluate many human optimization techniques and emerging wearable technologies, such as exoskeletons, physiological status monitors, and heads-up displays. The standard laboratory setting lacks many aspects of the real world that are often not considered during the design and development of technologies. A physically and cognitively immersive environment, such as the CAREN, provides a realistic testing platform for systematic and rapid evaluation even during early design phases while capturing many gold-standard laboratory metrics. Data from such testing will provide constructive feedback to developers and enrich the technology development process.

Operational Excellence

In challenging operational environments, field-site training is often difficult and expensive to develop, with limited opportunity for trainees to iterate through a variety of mission conditions and environments. The CAREN provides a rich and flexible serious gaming platform for operators to rapidly build experience by completing multiple scenarios in a single day. As a result, trainees, such as first responders and warfighters, can increase their physical agility, improve their mission-specific decision-making skills, and learn from their mistakes with minimal risk of injury.

Training for Success

Developing proficiency in a specialized physical or cognitive skill often requires years of experience for a person to become a qualified expert. The multimodal sensing capabilities of the CAREN can quantify specific physical characteristics that identify individuals as experts and use them as benchmarks for training. Providing biofeedback via sensory cues assists novices to more quickly improve their skills in tasks such as dismounted marksmanship or rapid threat identification. The CAREN provides a way to learn more about the basic physiology of skilled experts and an interactive method for enhanced learning.

About the Authors

Joseph J. Lacirignola is a member of the technical staff in the Counter–Weapons of Mass Destruction Systems Group. His research interests include sensor, device, and novel detection development capabilities. He joined the Laboratory in 2003 and worked on CANARY (i.e., Cellular Analysis and Notification of Antigen Risks and Yields) and related biosensing hardware platforms. Throughout his career, he has been involved in numerous hardware programs spanning biological and chemical detection, wearable sensing, novel 3D-printing hardware, multimodal detection systems, and the development of unmanned aerial and ground vehicles. He has authored several journal articles, conference presentations, and book chapters and holds more than 10 patents. He holds a bachelor's degree in biology from Syracuse University, bachelor's degrees in mechanical engineering and mechanical design from Wentworth Institute of Technology, and a master's degree in molecular and cell biology from Brandeis University.

James R. Williamson is a member of the technical staff in the Human Health and Performance Systems Group. His work focuses on detecting and modeling neurocognitive and physical states through the analysis of speech, facial expressions, gait, and body-sensing modalities. Prior to joining Lincoln Laboratory, he earned a bachelor's degree in psychology from the University of Massachusetts Amherst and a doctoral degree in cognitive and neural systems from Boston University. He also served as a research assistant professor at Boston University.

Shakti K. Davis is a member of the senior technical staff in the Human Health and Performance Systems Group. Among her research interests is the development of sensor and signal processing algorithms that are aimed at applications for protecting warfighters against injury and enhancing warfighters' performance capabilities. Since 2015, she has supported a variety of signal processing research efforts, including characterizing exposure to hazardous military noise through the use of body-worn acoustic sensors, developing algorithms to estimate risk factors of musculoskeletal injury by using foot-mounted biomechanical sensors, and investigating classification methods that use high-resolution noninvasive physiological sensors to provide early warning of virus exposure. She joined Lincoln Laboratory in 2006 and worked on detection and classification algorithms for airborne radar systems for nine years. She holds a bachelor's degree from New Mexico State University and a doctorate from the University of Wisconsin–Madison in electrical engineering.

Whitney T. Young was a member of the Bioengineering Systems and Technology Group for more than six years. While in the group, she contributed to a variety of projects, ranging from biodefense to synthetic biology to physiological monitoring. In her last few years at Lincoln Laboratory, she researched biomechanics, specifically data collection and analysis of gait in various loaded conditions, to detect and prevent musculoskeletal injuries in warfighters. She now works as a mission systems analyst at Ball Aerospace in Colorado. She holds a master's degree in mechanical engineering from MIT and a bachelor's degree in bioengineering from the University of Toledo.

The Impact of Host-Based Early Warning on Disease Outbreaks

Mark A. Hernandez, Lauren E. Milechin, Shakti K. Davis, Richard A. DeLaura,

Kajal T. Claypool, and Albert J. Swiston

Lincoln Laboratory researchers investigated how early warning of exposure to pathogens could shape health care responses to disease outbreaks. Basing their analysis on the capability of an innovative algorithm that enables the detection of pathogen exposure in individuals before symptoms of disease occur, the research team evaluated the effects of such early warning on various strategies for mitigating a widespread outbreak. While not specifically focused on COVID-19, this effort lays the foundation for understanding the impact of early warning technology in combination with other nonpharmaceutical interventions.

>> **The detection of communicable** pathogens responsible for major outbreaks often relies on health care professionals' recognition of symptoms manifesting in infectious individuals. Early warning of such communicable diseases before the onset of symptoms could improve both patient care and public health responses. MIT Lincoln Laboratory's PRESAGED (Presymptomatic Agent Exposure Detection) algorithm uses host-based physiological signals to detect an individual's exposure to pathogens, such as viruses and bacteria, before overt symptoms emerge and infectiousness is peak.

Researchers in the PRESAGED program have used data from non-human primate studies to demonstrate that the algorithm can provide two to three days of early warning before the onset of incipient symptoms (e.g., fever), independent of the particular pathogen, exposure route, pathogen dose, or animal species [1]. These results are consistent with recent findings of Speranza et al. that show presymptomatic upregulation in biomarkers potentially linked to pathogen exposure around the same time in non-human primates exposed to Ebola [2].

The PRESAGED algorithm relies on data that could be collected noninvasively, such as heart electrical activity (electrocardiography, or ECG), to predict the probability that an individual has been exposed to a pathogen. This system addresses fundamental limitations of existing highly sensitive and specific bioassays: specifically, that these tests require the collection of blood or other biofluids and that samples are only collected and analyzed when there is a high suspicion of exposure. Furthermore, most assays focus on detecting proteomic or genetic signatures

of a pathogen itself, at which time an individual will likely already have symptoms and thus be infectious [3].

PRESAGED offers potential applications in health care responses to emerging pathogens on two fronts. The first would be improved individual patient care; medical countermeasures are nearly always more effective when deployed before overt indications of infection [4–6]. The second would be faster and more confident implementation of public health measures to mitigate the spread of disease outbreaks. Lincoln Laboratory researchers have been focusing on the latter, investigating how host-based early warning capabilities could lead to effective pathogen surveillance systems and public health interventions. Although much work has focused on predicting disease outbreaks based on environmental and vector-related data, such as weather patterns, geography, and vegetation conditions [7, 8], less work has explored using the conditions of the host itself for predictive purposes.

This research is of particular interest to the Department of Defense because U.S. servicemembers deployed worldwide can potentially come in contact with a huge variety of communicable diseases, ranging from rarely encountered Ebola to annual influenza. Previously unknown emerging pathogens, such as Chikungunya and Zika, have the potential to disproportionally affect military personnel who are often deployed to regions where such pathogens are prevalent.

Numerous papers have applied epidemiological models to characterize the disease transmission dynamics and the effectiveness of public health interventions of past outbreaks [9–13]. A few papers have incorporated a hypothetical early diagnostic capability (based on advancements in bioassay tests) in their models and evaluated its effectiveness in controlling the spread of disease outbreaks [14]. However, there has yet to be an effort that assesses the potential impact of a host-based early warning system.

We have developed a series of epidemiological models that quantify the potential impact that a host-based early warning capability would have in mitigating pathogen transmission during an outbreak. Our epidemiological models reflect a variety of traditional public health policies related to nonpharmaceutical interventions, such as quarantine and patient isolation, and novel policies enabled by host-based early warning capabilities. For each policy-dependent model, we simulated outbreak scenarios

and calculated the size of the outbreak (total number of infections) and the operational burden (total number of lost duty days resulting from quarantine or isolation). These metrics were used to understand the trade space for the different policies. Our simulations demonstrate the utility of host-based early warning systems in controlling an outbreak under various outbreak conditions, further motivating discussion about the potential benefits and limitations of population-wide implementation in the U.S. armed forces or the broader civilian population.

Models for Disease Outbreak Simulation
The SEIR Model

One of the most common epidemiological models for simulating disease outbreak scenarios is the deterministic SEIR model. This approach splits a given population into separate compartments defined by their relationship to a disease outbreak [15]:

1. Susceptible: healthy individuals who can be exposed to the pathogen
2. Exposed: individuals who are in the incubation phase; they have been exposed to the pathogen but are not yet showing symptoms and are not infectious
3. Infectious: individuals who are infectious to the susceptible population and will eventually display overt symptoms
4. Recovered: those who have recovered from illness and acquired immunity to further infection

These compartments are then linked with a system of ordinary differential equations (ODEs) to characterize how individuals transition into and out of each compartment over time. A variety of scenarios can be simulated by changing the rate parameters of the ODEs linking the population compartments. Furthermore, this approach allows quantitative projections of how many people are exposed to a pathogen and become sick under different outbreak conditions.

While the SEIR model is mathematically rigorous and often has good predictive utility during an outbreak, the SEIR model requires several assumptions. First, the model assumes there is a fixed population N, with no births or deaths other than those resulting from the infectious disease. The next assumption is that the population is homogeneously mixing, meaning that transmission between any two individuals is equally likely. The model also assumes that exposed individuals become infectious

after a fixed incubation period, thus not accounting for individual variability of disease progression for the young, elderly, or immunocompromised. Finally, it is assumed that all recovered individuals are immune to further infection and thus do not re-enter the susceptible class. While any of these assumptions may fail to hold in particular contexts, abiding by them allows for greater mathematical tractability and offers similar relative output trends.

Baseline SEIR Model

The SEIR model consists of a system of four ODEs that describe the rate of change of individuals in each compartment over time (Figure 1). The differential equations for Susceptible, S_t, Exposed, E_t, Infectious, I_t, and Recovered, R_t, are as follows, where dependence on time, t, is omitted in the notation below for simplicity:

$$\frac{dS}{dt} = -\frac{\beta SI}{N}$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E$$

$$\frac{dI}{dt} = \sigma E - \gamma I$$

$$\frac{dR}{dt} = (1 - f)\gamma I$$

Table 1 describes all of the parameters we use for our models. At the start of the model, a subset of the population is exposed to a pathogen (E_0), and the remainder are susceptible to infection. Susceptible individuals enter the exposed compartment at a rate of $\beta I/N$, which is known as the normalized transmission rate. The parameter β is the contact rate, which accounts for how often susceptible-infectious contacts result in a susceptible individual becoming exposed to the pathogen. Exposed individuals become infectious at a rate of σ, which is the inverse of the incubation period. Infectious individuals stay infectious at a rate of γ (the inverse of the infectious period) until they recover or die. We define the recovery rate, ρ, as $(1 - f)\gamma$ where the case fatality rate f is the proportion of infected individuals who die from the disease. Note that we do not specify a mortality compartment in this analysis, though for some pathogens spread through the handling of remains (such as Ebola during the 2014 West Africa outbreak), this compartment would be a critical addition.

Another important value given by this model is the basic reproduction number $\mathcal{R}_0$, which is often included as a basic property of any given pathogen [15]. $\mathcal{R}_0$ represents the average number of additional infections caused by each infectious individual, assuming there are no control interventions. In a fixed population represented in the SEIR model, the reproduction number can be calculated from parameters in the model as $\mathcal{R}_0 = \beta/\gamma$.

The solutions of the SEIR model are functions of the number of individuals in each compartment with respect to time, i.e., S_t, E_t, I_t, and R_t. We solve this numerically by using an ODE solver in MATLAB®, and with these results, we may now answer some of the initial questions posed. For example, if we want to know the number of people who are not showing symptoms (visibly healthy individuals), we could plot $S + E + R$.

An example output of the full SEIR model is shown in Figure 2a. Here, we begin with a population of $N = 1000$ people (roughly equivalent to a large battalion), and consider a scenario in which 300 individuals are exposed to some pathogen at time, $t = 0$ ($S_{t=0} = 700, E_{t=0} = 300, I_{t=0} = R_{t=0} = 0$). If we assume a particular infectious pathogen has a contact rate, $\beta = 0.75$, with an incubation period of two days, $\sigma = 0.5$, and an infectious period of five days, $\gamma = 0.2$ (similar properties to a highly contagious flu virus), the solution to the SEIR model shows that nearly the entire population will contract the disease over the course of a month. Figure 2b focuses on the number of cumulative infections, a metric for assessing the overall size of the outbreak over time. This baseline condition, in which no public health policy is in effect, represents the most severe outcome from an infectious disease outbreak.

Policy-Dependent SEIR Models

In the previous section, we considered the standard SEIR model, which describes disease spread *without* any public health interventions. To quantitatively evaluate an outbreak scenario with additional measures, the parameters and compartments of the SEIR model may be modified to reflect policy choices or new early warning technologies. Two policies currently considered as standards in handling possible outbreaks rely on self-monitoring to implement voluntary isolation or quarantining all individuals that may have been exposed

TABLE 1. Parameter Definitions and Values for Baseline and Policy-Dependent Models

PARAMETER	DEFINITION	VALUE FOR THIS ANALYSIS
N	Total population	1000
E_0	Initially exposed population	300
β	Contact rate	0.75
σ	Incubation rate	0.5 (2-day incubation period)
Y	Infectious rate	0.2 (5-day infectious period)
ρ	Recovery rate	0.18
f	Case fatality rate	0.1
θ	Self-reporting probability	1 (all ill individuals self-report)
α	Self-reporting rate	0.5 (2-day self-report) delay
T_Q	Time of quarantine	0
ε	Release rate	0.2 (5-day quarantine period)
κ	False-alarm rate	1/365 (1 false alarm per year)
λ	Early detection rate	1 (1-day early detection time)
P_D	Probability of early detection	0.8
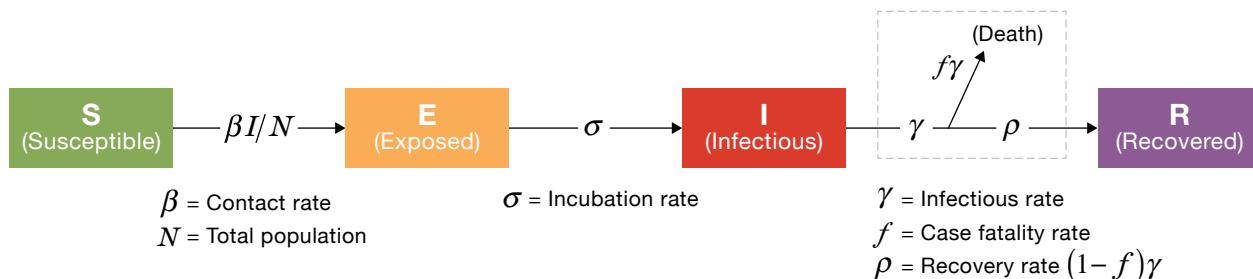
FIGURE 1. In this schematic diagram of the SEIR model, each box represents a compartment of the population, and arrows show the progression of individuals through those compartments. Expressions above arrows are rate coefficients showing progression through compartments. Note that the dashed boxed portion, indicating individuals leaving the population because of death, will not be included in future diagrams for simplicity.

to a pathogen. Isolation applies to individuals who are already ill, whereas quarantine applies to individuals who may have been exposed but have not shown symptoms. The first policy, self-monitoring and reporting, which assumes individuals self-report when they develop symptoms, allows for additional infections during the time delay between the onset of symptoms and isolation; however, it does not require the logistical and financial cost of quarantine. The second policy, quarantining the entire population, is effective in reducing additional transmissions but requires prohibitively onerous costs for large populations. In an effort to explore an alternative to these two policies, we modified our SEIR model to simulate a third policy, quarantine-on-alert, in which

MARK A. HERNANDEZ, LAUREN E. MILECHIN, SHAKTI K. DAVIS, RICHARD A. DELAURA, KAJAL T. CLAYPOOL, AND ALBERT J. SWISTON
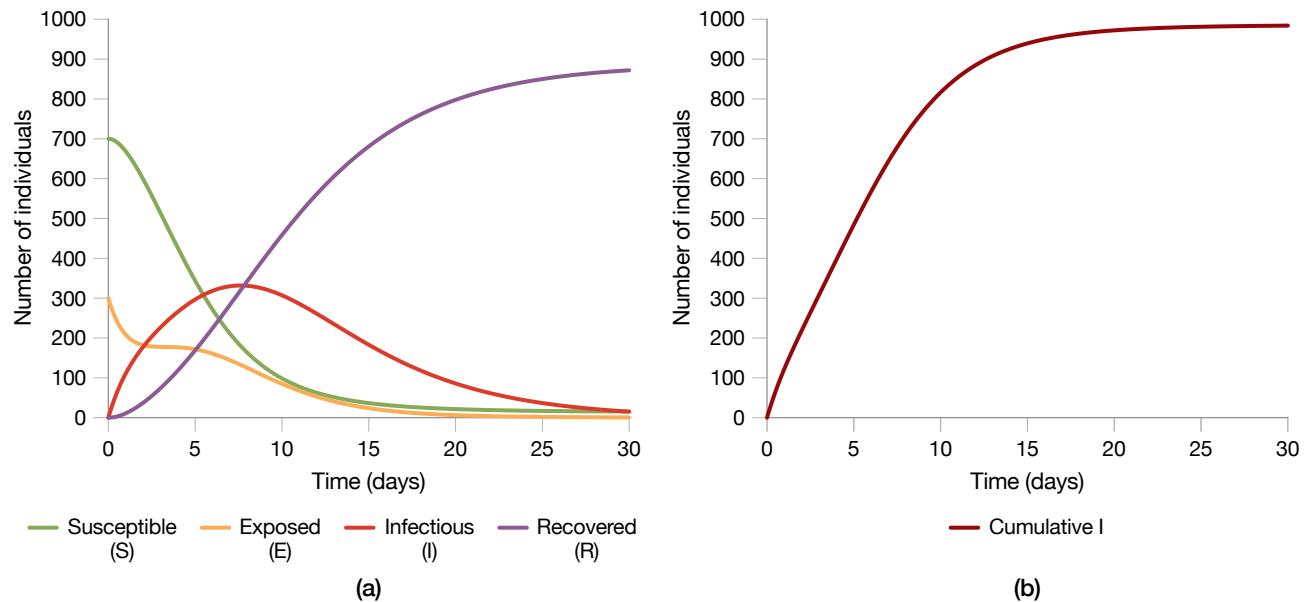
FIGURE 2. A numeric solution for the baseline SEIR model shows the population of each compartment versus time (a), which indicates that for an initial exposure of 300 individuals, nearly the entire population will eventually fall ill; the cumulative number of infections (Cumulative I) versus time is shown in (b).
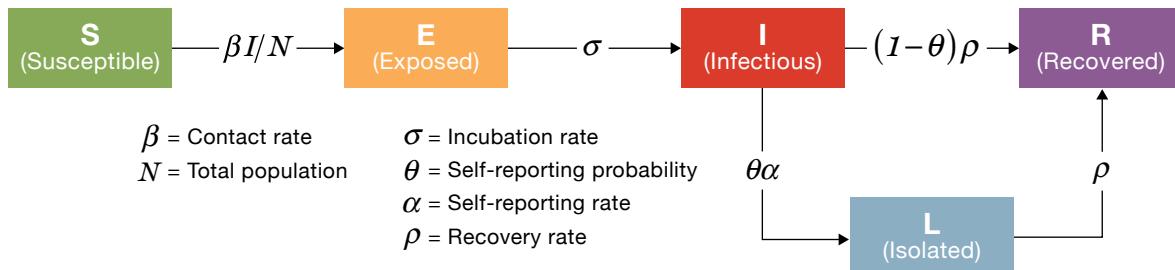
FIGURE 3. The model diagram for a self-monitoring and self-reporting policy includes an additional compartment (*L*) for individuals who are isolated after developing symptoms. Note that fatalities are left out for simplicity.

individuals are only quarantined when prompted by a host-based early warning system. We hypothesized that the early notification of incipient illness will allow for reductions in pathogen transmission while minimizing the number of quarantined individuals. We then assessed each of these three new policy-dependent SEIR models.

Self-Monitoring and Reporting: Isolation after Symptoms

In the course of an infectious disease outbreak, most individuals will self-monitor for symptoms of the pathogen. If individuals start to develop symptoms, they can self-report to a medical facility and may be

immediately isolated until they recover. A new compartment, isoLated (*L*), was added to the model to reflect this symptomatic and infectious population, which has limited contact with the rest of the population and therefore reduced transmission potential (Figure 3). The self-reporting probability θ is defined as the proportion of symptomatic individuals who are compliant in reporting their symptoms and enter the *L* compartment, while the self-reporting rate α is derived from the delay between developing symptoms and self-reporting (i.e., the inverse of the average time delay). For our model, we assumed that isolation is 100 percent effective in preventing transmission, and individuals in the *L* compartment
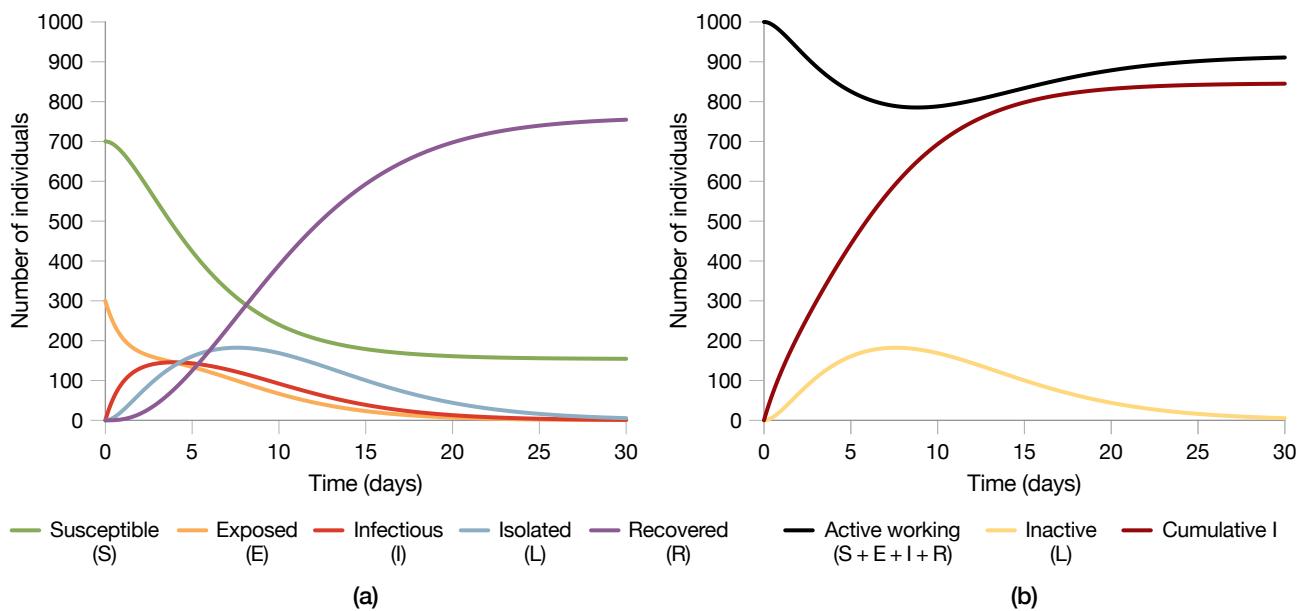
FIGURE 4. In this numeric solution for our SEIR model with a self-monitoring and isolation policy enacted, the population of each compartment versus time is shown in (a), and the active working ($S + E + I + R$) and inactive (L) populations versus time with the cumulative infections versus time are shown in (b). The self-monitoring policy has effectively reduced the total number of disease cases and blunted the outbreak's peak severity (number of cases in I and L peak at approximately 7 days).

are unable to infect the susceptible population. While a self-monitoring policy reduces transmission rates by isolating the self-reported sick, both the latency α and an imperfect self-reporting probability ($\theta < 1$) will still lead to opportunities for transmission.

Outputs for the SEIR model with self-reporting and isolation are shown in Figure 4. This policy addition leads to both fewer total individuals contracting the disease and fewer individuals acutely symptomatic at the height of the outbreak. The new isolated compartment L, however, has other costs associated with lost duty days, mandated isolation, medical facilities, and treatment. The choices of self-reporting probability and rate are critical in this scenario, and, in reality, may have such broad distributions that relying on this policy in an acute outbreak may do little to prevent additional infections.

Quarantining the Entire Population

Another possible, though draconian, public health response to an outbreak is to quarantine currently healthy individuals with some likelihood of pathogen exposure. In outbreaks with a large population subset having some exposure likelihood, a quarantine-all policy would separate all individuals in the population from

contact with one another. Our model assumes that quarantined individuals, represented by the Q compartment, are monitored and immediately isolated once they develop symptoms (Figure 5). If quarantined individuals do not become symptomatic after the maximum incubation time of the suspected pathogen (often several weeks), they are released and re-enter the S compartment at a rate of ε (the inverse of the maximum incubation time). As shown in Figure 6, the quarantine-all policy eliminates opportunities for further pathogen transmission; however, it also results in the quarantine of healthy individuals who have not been exposed, contributing to immense productivity losses, extreme logistic burdens associated with providing accommodations for an entire population, and acute civil rights concerns. Furthermore, the assumption that transmission is zero within the Q compartment may not be realistic because transmission for some pathogen may occur before overt symptoms and isolation. This quarantine-all policy could result in the illness of people who, if not for the quarantine, would never have been exposed to the pathogen. Blanket quarantine policies are extreme examples of public health interventions and represent a clear scenario that new technologies may improve.

MARK A. HERNANDEZ, LAUREN E. MILECHIN, SHAKTI K. DAVIS, RICHARD A. DELAURA, KAJAL T. CLAYPOOL, AND ALBERT J. SWISTON
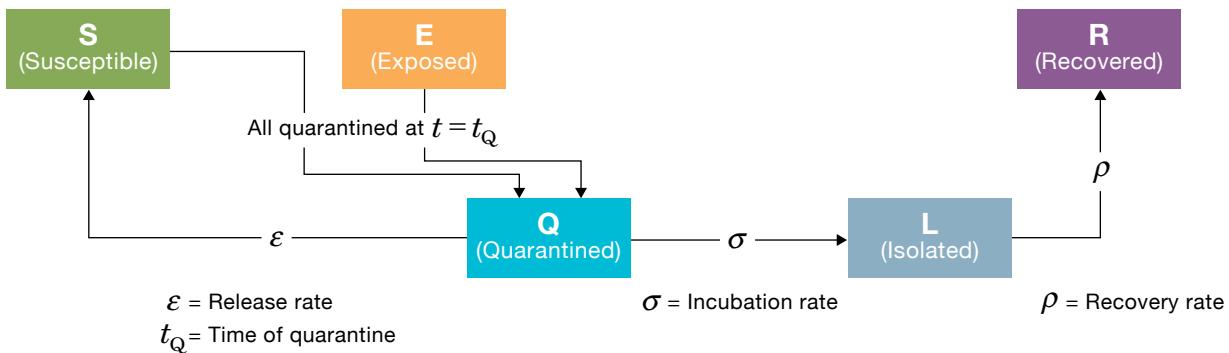
FIGURE 5. The model diagram for a quarantine-all policy moves all susceptible and exposed individuals to the quarantined compartment (Q) at the time $t_Q = 0$ days. Quarantined individuals who do not develop symptoms are returned to the susceptible compartment after a maximum incubation time.
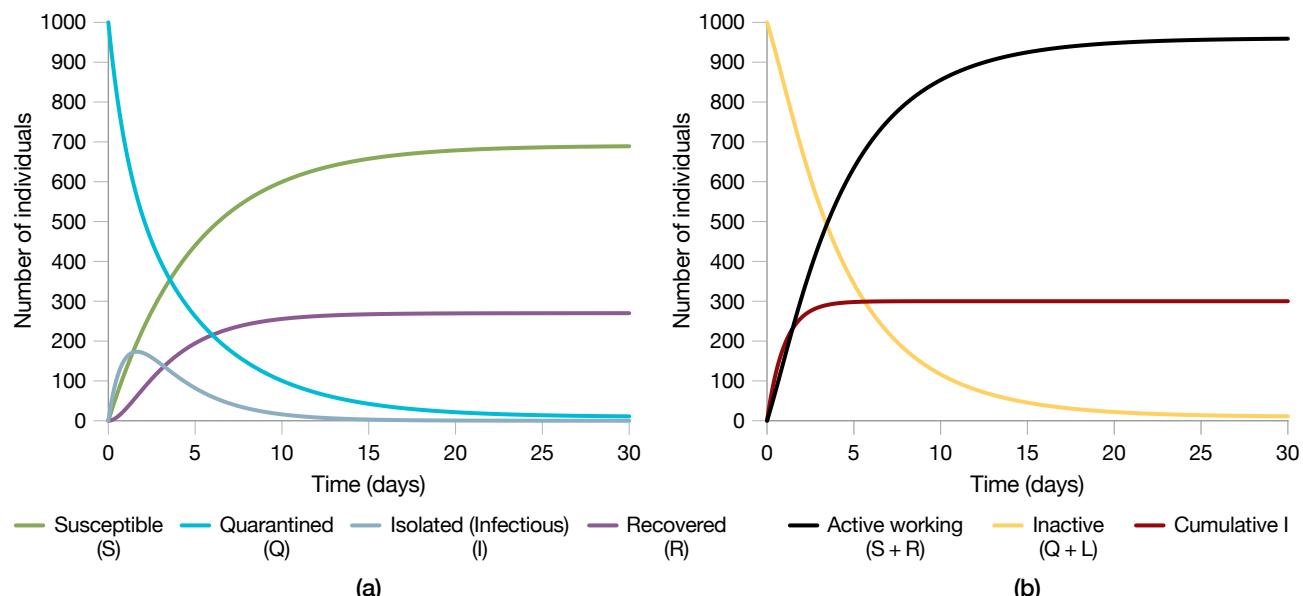
FIGURE 6. In the numeric solution for our SEIR model with a quarantine-all policy enacted, the population of each compartment versus time, which notably does not include an exposed compartment, is shown in (a). In (a) are individuals suspected to be exposed and quarantined initially, until they either fall ill (and are isolated) or are released after the quarantine duration (21 days). The active working (S + R) and inactive (Q + L) populations versus time with the cumulative infections versus time are shown in (b). The quarantine-all policy has very effectively reduced the total number of disease cases but has vastly increased the burdens of quarantine and isolation.

Quarantine on Alert

The third model we explored simulates a novel policy that could potentially be enabled by future host-based early detection capabilities. Here, a host-based early warning system could prompt a quarantine only on alert when the system detects presymptomatic signs of exposure. We integrated the early warning performance metrics from our PRESAGED algorithm—population-wide probability of early detection, daily false-alarm rate, and early

detection time—into a base SEIR model. As shown in Figure 7, a host-based early warning system would trigger exposed individuals to transition into quarantine at a rate characterized by the product of the parameters λ and P_D. The parameter λ is an early detection rate defined as the inverse of the system's average early detection time for a given pathogen. The parameter P_D is the population-wide probability of early detection, i.e., the fraction of the exposed population that will present early detections of

β = Contact rate
N = Total population
κ = False-alarm rate
ε = Release rate

P_D = Probability of early detection
σ = Incubation rate
λ = Early detection rate

ρ = Recovery rate
θ = Self-reporting probability
α = Self-reporting rate
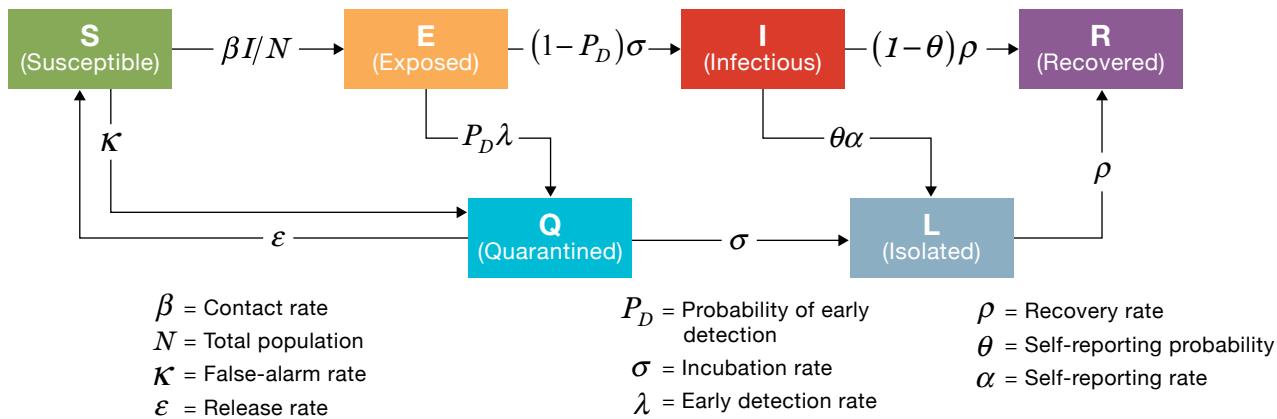
FIGURE 7. This hybrid model diagram illustrates the final quarantine-on-alert policy. The performance of the PRESAGED-like early warning system is now explicitly added, including the probabilities of early detection and false alarm, as well as an expected early detection time before symptom onset and infectiousness. This hybrid between quarantine-all and self-monitoring with isolation seeks to leverage the epidemic-limiting behavior of both policies while reducing the cost and burdens of each.

pathogen exposure prior to becoming infectious. Missed detection cases, when the system fails to produce an early warning, occur with a probability of $1 - P_D$ across the exposed population. For this policy, false alarms of the host-based early warning system would cause healthy individuals to be incorrectly quarantined. Our model takes this notion into account by adding a transition rate from the susceptible to the quarantined compartment, characterized by a daily false-alarm rate κ, the probability of an individual's early warning system presenting a false alarm on a given day.

While the self-monitoring and quarantine-all policies present clear trade-offs between cumulative number of infections and number of people isolated or quarantined, as shown in Figure 8, this quarantine-on-alert policy offers a hybrid approach that could optimize for both. Quarantine-on-alert vastly reduces quarantine costs by selectively identifying likely exposed individuals through physiologically based predictions. Furthermore, once quarantined, these high-risk individuals can be immediately isolated, thus limiting the opportunities of additional infections and enabling more focused medical care.

SEIR Model Reproduction Numbers by Policy Choice

Modifications made to the baseline SEIR model result in changes to the reproduction number ($\mathcal{R}$) associated with each policy-dependent SEIR model [16]. Our example scenario assumes that every symptomatic person will

self-report and thus the self-reporting probability, θ, is 1. Table 2 summarizes equations to calculate $\mathcal{R}$ values for each policy, the $\mathcal{R}$ values given the parameter values in Table 1, and the cumulative cases of infection for each policy choice after 50 days. The policies with the lower $\mathcal{R}$ values (quarantine-all and quarantine-on-alert) are associated with minimal pathogen spread, while the policies with higher $\mathcal{R}$ values (self-monitoring and baseline) yield a greater number of infection cases. The number of cumulative infections as the ODE reaches equilibrium illustrates how the reproduction number would affect the final size of the outbreak. Note that the number of cumulative infections for baseline saturates around the population size ($N = 1000$) when the number of individuals in the susceptible compartment is exhausted.

Policy Trade-Space Analysis

The policy-dependent SEIR models provide a foundation for a trade-space analysis of different quarantine, isolation, and treatment (QIT) policies as a function of disease transmission characteristics, exposure scenarios, and performance of early warning systems. The analysis captures both the potential benefits (reduction in quarantine costs, more focused use of medical resources) and risks (increase in infections) for each policy.

To demonstrate the potential utility of the early warning–enabled quarantine-on-alert model in a more comprehensive QIT policy analysis, we compare outcomes over a range of disease transmission rates,
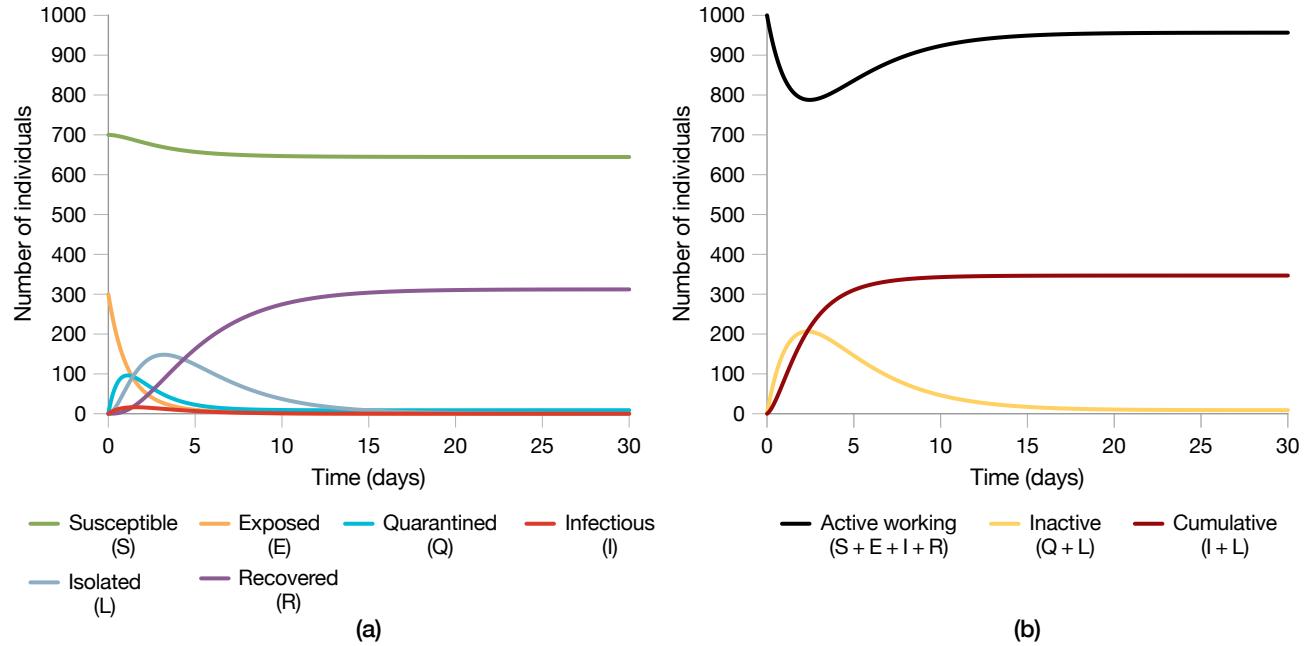
FIGURE 8. This numeric solution for our SEIR model with a quarantine-on-alert policy enacted shows the population of each compartment versus time (a); the plot shows how such an early warning technology could limit the outbreak size. As the initially exposed individuals receive alerts, they enter the quarantine-on-alert compartment and are unable to infect the susceptible compartment. The active working ($S + E + I + R$) and inactive ($Q + L$) populations versus time are shown with the cumulative infections versus time in (b). The quarantine-on-alert policy has effectively reduced the additional disease cases and has largely limited those in quarantine to those individuals who were initially exposed.

initial exposure conditions, and early warning performance parameters. We vary these parameters in simulations of a mass exposure to an infectious pathogen occurring in a population of 1,000 people. All scenarios were simulated over 50 days. Because we fix the exposure time for all cases to $t = 0$, rather than using the quarantine release rate ε, all individuals in quarantine are returned to the susceptible compartment at $t = 21$ days (the maximum incubation period).

Policy outcomes are defined by two metrics that are derived from the policy-dependent SEIR model outputs as shown in Figure 9. The first, lost duty days, is expressed as a percentage of total number of days of work productivity that are lost because of quarantine and isolation in the 50-day simulation; this percentage is proportional to the integral of the curves in Figure 9a. The second metric, cumulative infections, is the percentage of the population that has been infected by the end of the simulation (Figure 9b). These two metrics characterize a trade space for evaluating QIT policies under different circumstances, because any measure that reduces the number of people

in quarantine and isolation may be expected to increase the likelihood of infection in the population. While these metrics are helpful for evaluating the impact of various policy choices, they are by no means comprehensive and particularly do not consider the financial or other costs associated with quarantine.

Five QIT policy sets were evaluated: quarantine-all, isolate upon self-reporting, and quarantine-on-alert with three different levels of early detection performance—high sensitivity ($P_D = 0.8$, $\kappa = 0.10$), high specificity ($P_D = 0.4$, $\kappa = 1/365$), and a near-ideal early-detection system ($P_D = 0.8$, $\kappa = 1/365$). These five QIT policy options were then tested against four different outbreak scenarios with high or low disease transmission (where $\beta = 0.6$ or 0.3, respectively) and high or low initial population exposures (where $E_0 = 600$ or 50, respectively). Figure 10 shows the QIT policy trade-offs for these 20 independent outbreak and policy combinations.

For all scenarios, a quarantine-on-alert policy coupled with the near-ideal early warning capability reduces quarantine needs with only a small increase in

TABLE 2. Reproduction Number ($\mathcal{R}$) Equations and Nominal $\mathcal{R}$ Values for Our Example Scenario

POLICY	$\mathcal{R}$ EQUATION	$\mathcal{R}$ VALUE	CUMULATIVE INFECTIONS AFTER 50 DAYS
Baseline	$\mathcal{R}_0 = \dfrac{\beta}{\gamma}$	3.75	986
Self-monitoring	$\mathcal{R} = \dfrac{\beta}{\alpha}$	1.5	846
Quarantine-all	$\mathcal{R} = 0$	0	300
Quarantine-on-alert	$\mathcal{R} = \dfrac{\beta}{\alpha} \dfrac{(1-P_D)\sigma}{(1-P_D)\sigma + P_D \lambda}$	0.167	347

In a population of $N = 1000$, the number of initially exposed individuals is $E_0 = 300$. The nominal R Value and Cumulative Infections were calculated using parameter values from Table 1.

the number of additional infections. The cost of focusing on a high-specificity early detection system (i.e., a reduction in false alarms and thus quarantine costs) is an increase in additional infections relative to the near-ideal system. Conversely, a high-sensitivity system increases the percentage of the population in quarantine compared to both the ideal and high-specificity early detection system while also reducing the number of additional infections to nearly the number seen by quarantining the entire population a priori.

For scenarios with low initial exposure, the impact of early warning–enabled QIT policies varies dramatically with transmissibility. For low-transmissibility and small initial exposures (Figure 11a), an early warning capability demonstrates the least utility because the outbreak seems to be effectively contained with just self-monitoring. However, for a pathogen with high transmissibility (Figure 11b), the cumulative infections are reduced significantly with any of the notional early warning systems. A high-specificity system offers a particularly promising result, showing comparable infection reduction to the near-ideal system while also minimizing the number of days lost to quarantine.

In the scenarios of high initial exposure (Figure 11c and Figure 11d), the costs of imperfect early detection are more pronounced compared to those for scenarios of low initial exposure. Because a high initial exposure greatly increases the likelihood of infectious individuals

transmitting the pathogen to those susceptible, even a short delay in isolating infectious individuals will lead to more infections. The reduction of cumulative infections for all notional early warning systems, relative to the self-monitoring policy, is less prominent than in scenarios of low initial exposure because the initial exposed population is close to saturation. In effect, for these extreme cases of mass exposures, the more aggressive policies (i.e., a high-sensitivity early detection system or quarantining everyone) may be more effective.

Discussion and Future Work

The results for the early warning–enabled and policy-dependent SEIR models allowed for a quantitative analysis of the QIT trade space and provided guidance on priorities for the future development of early warning technology. However, this illustration of QIT risk analysis captures only a subset of the factors that must be considered in the formulation of a rational, effective QIT policy. The context of the scenario will ultimately inform where the ideal operating point would be. For example, the number of total infections may be interpreted differently depending upon the virulence of the disease. The tolerance for new infections may be low if the consequences are high, such as if the infection is almost always fatal or is associated with severe symptoms and long-term complications. Additionally, the availability of diagnostic tests may further refine the use of early warning capabilities;
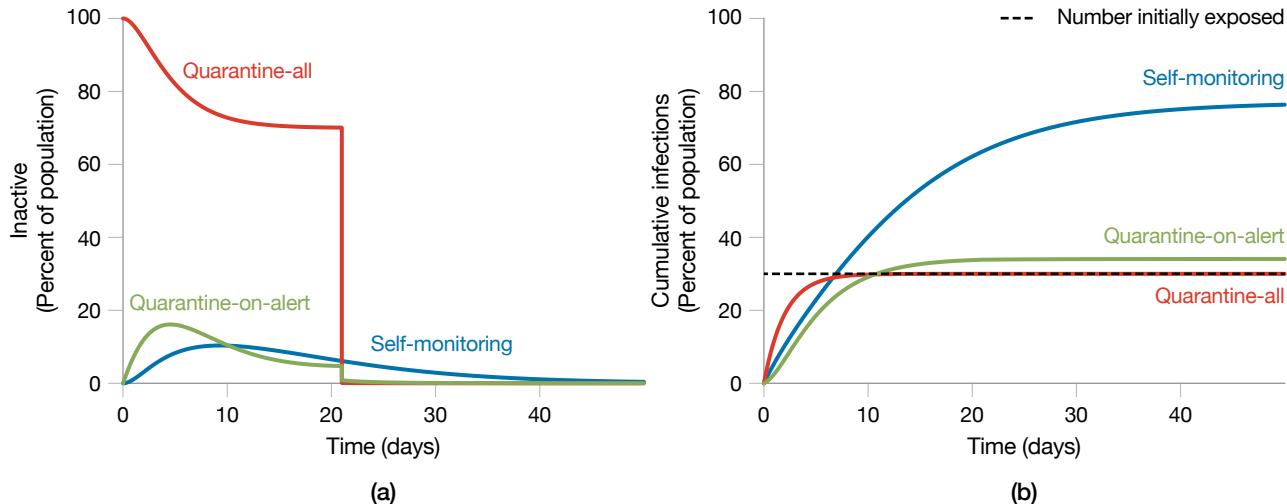
FIGURE 9. The model outputs compare consequences of self-monitoring, quarantine-all, and quarantine-on-alert outbreak response policies. In (a), the plot shows the number of inactive individuals who are unable to work because of quarantine and isolation ($Q + L$ for quarantine-all and quarantine-on-alert, only L for self-monitoring because it does not utilize quarantine) versus time. The sharp drop-off for quarantine-all represents the release of all quarantined individuals after 21 days. In (b), the plot shows the cumulative infections versus time. Any infections above the dotted line (number of individuals initially exposed at $t = 0$) indicate additional transmission of the pathogen.

a more sensitive early detection capability could be combined with a cued use of diagnostic testing to form a much more targeted approach, reducing costs and the likelihood of a false detection. While the results we presented are a first attempt at understanding the potential utility of a host-based early warning system during an outbreak, a comprehensive risk assessment of a QIT policy must consider a range of factors:

- Disease characteristics: prevalence, transmissibility, incubation period, and severity
- Response options: reliable diagnostic tests, vaccines, or treatments
- Resources: cost and availability of QIT measures
- Early warning capability: performance characteristics of early warning systems
- Public health infrastructure: ability to implement an effective public health campaign

An element that we do not include in this model is patient care measures, which are nearly always more effective when deployed earlier than overt indications of infection (e.g., fever). For instance, antiviral drugs (such as zanamivir and oseltamivir/Tamiflu) are most effective in the first ~48 hours of symptoms [8, 17, 18]; PRESAGED-enabled early warning would allow much faster prescription, use, and potentially more profound
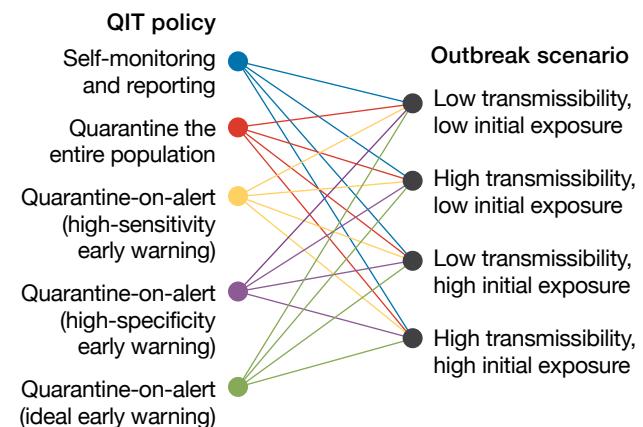
FIGURE 10. Simulations are run on each combination of QIT policy and outbreak scenario, resulting in 20 independent conditions for the trade-space simulations.

therapeutic impact for current dosage recommendations. Triggering the use of diagnostics early would allow clinicians to target drug-based interventions, such as using the precise antibiotic for bacterial infections rather than relying on broad-spectrum versions that contribute to the evolution of drug-resistant bacteria ("superbugs"). Finally, simple supportive care for mild infections, e.g., having rehydrating solutions or over-the-counter
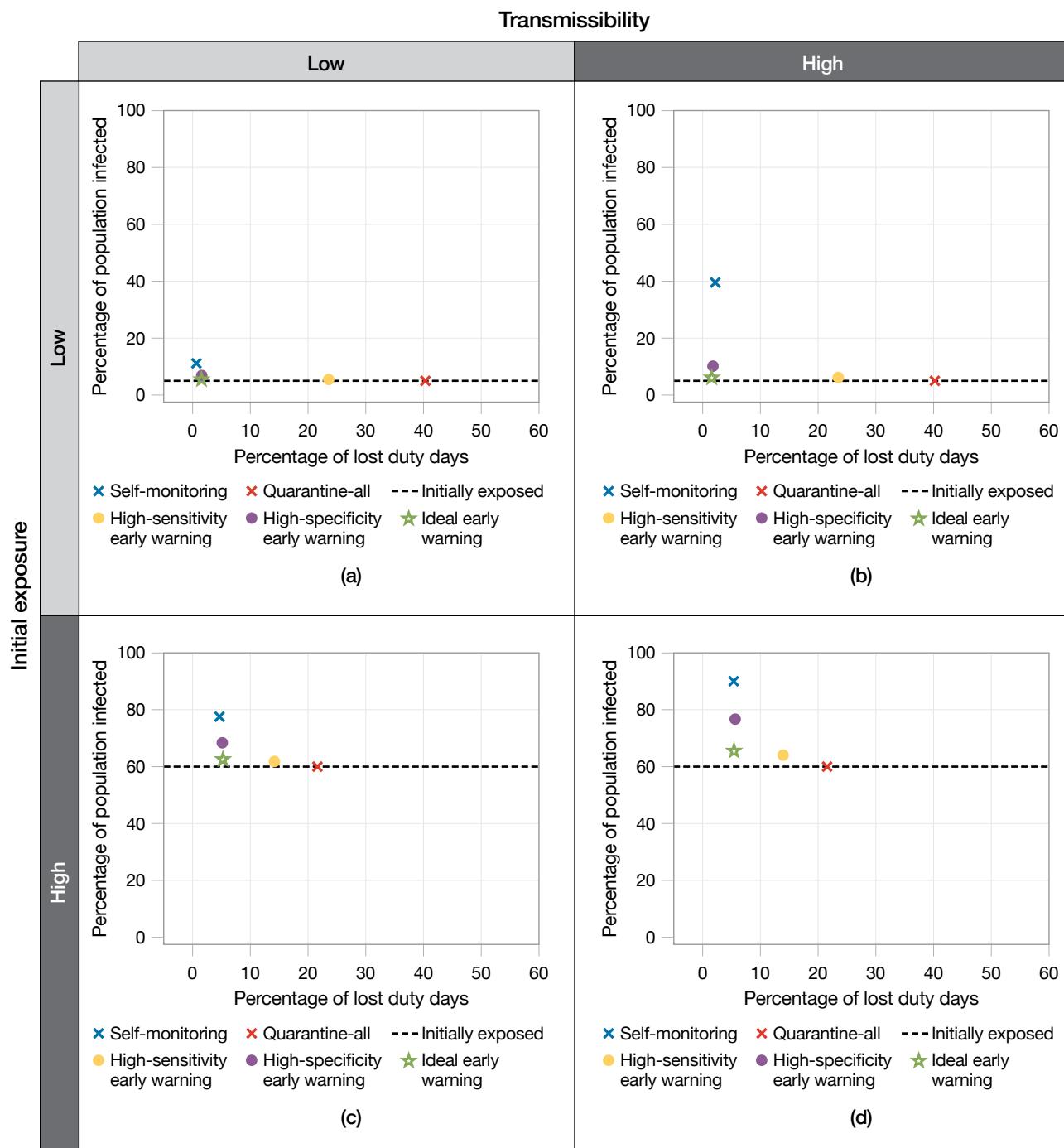
FIGURE 11. The trade space for the percentage of population infected versus percentage of lost duty days compares the augmented policies: self-monitoring, quarantine-all, and quarantine-on-alert with three levels of early warning performance (high sensitivity, high specificity, and ideal early warning). Four situations were considered: low-transmissibility pathogen with a low initial exposure (a), high-transmissibility pathogen with a low initial exposure (b), low-transmissibility pathogen with a high initial exposure (c), and high-transmissibility pathogen with a high initial exposure (d).

MARK A. HERNANDEZ, LAUREN E. MILECHIN, SHAKTI K. DAVIS, RICHARD A. DELAURA, KAJAL T. CLAYPOOL, AND ALBERT J. SWISTON

symptom-relieving medications available, could ease the burden on health care workers in hospitals, nursing homes, or college campuses. All of these exciting possibilities could be enabled or improved with earlier detection.

Future modeling efforts will focus on improvements to this model, especially as more is known about the mechanistic basis for host-based early warning of pathogen exposure. Several additional aspects of QIT policy that will be considered include the following:

1. Consequences of infection, or a cost function of being in the infectious compartment. Adding a more explicit fatality compartment is straightforward. However, more complex models incorporating significantly time-delayed or nonlinear costs are much more accurate for measuring the effectiveness of QIT policies.

2. Early quarantine release. When the maximum incubation period may result in long quarantine periods that lead to unacceptably high quarantine costs, the absence of an exposure detection in individuals who have not been exposed to the pathogen may trigger an early release from quarantine. This early quarantine release approach could significantly reduce costs and civil rights issues associated with quarantine.

3. Modeling of additional QIT responses. Incorporating diagnostic tests into the models could refine the quarantine trigger and release policies. Additionally, future models could include specific treatments upon early warning to modulate infection outcomes. Such options may be particularly important in circumstances when quarantine and isolation resources are limited.

4. Movement among populations. The policy-dependent SEIR model addresses a single, isolated population being homogeneously mixed. However, in many circumstances, QIT policy is driven by a concern for the spread of the pathogen into connected populations. To address these circumstances, the policy-dependent SEIR model may be further extended to incorporate changes in each population compartment that result from the movement of people into (or out of) individual population compartments from (or to) other populations. In this formulation, individual populations are modeled with the extended SEIR mode as nodes in a migration network. Network edges are characterized by migration rates between the connected subpopulations.

5. Stochastic modeling. In a more realistic model, each of the SEIR parameters may be modeled in greater detail as a probability distribution rather than a fixed value. Stochastic modeling would support a more comprehensive risk analysis that includes insights about the likely range of potential outcomes, as well as rare but high consequence "edge" cases.

In conclusion, we have shown the epidemiological value of host-based early warning systems in a variety of pathogen outbreaks. By adjusting the underlying assumptions, both of the outbreak and the system performance metrics of a notional early detection system, we show in which scenarios early detection is most impactful. The results of this work emphasize the value of early detection in modulating public health responses, though future efforts will also include the value to individual patients. Current efforts at Lincoln Laboratory are focusing on the ability to monitor pathogen exposure of annual influenza, the ability to measure pathogen transmissibility noninvasively, and the impact this detection capability will have on patient care, public health responses, and service member readiness. ■

References

1. L. Milechin, S. Davis, T. Patel, M. Hernandez, G. Ciccarelli, S. Schwartz, et al., "Detecting Pathogen Exposure during the Non-symptomatic Incubation Period Using Physiological Data," bioRxiv: 218818-218818, 2017.

2. E. Speranza, S.L. Bixler, L.A. Altamura, C.E. Arnold, W.D. Pratt, C. Taylor-Howell, et al., "A Conserved Transcriptional Response to Intranasal Ebola Virus Exposure in Nonhuman Primates Prior to Onset of Fever," *Science Translational Medicine*, vol. 10, no. 434, 2018.

3. R. Liu, X. Wang, K. Aihara, and L. Chen, "Early Diagnosis of Complex Diseases by Molecular Biomarkers, Network Biomarkers, and Dynamical Network Biomarkers," *Medicinal Research Reviews*, vol. 34, no. 3, 2014, pp. 455–478.

4. P.K. Tosh and P. Sampathkumar, "What Clinicians Should Know about the 2014 Ebola Outbreak," *Mayo Clinic Proceedings*, vol. 89, no. 12, 2014, pp. 1710–1717.

5. D.G. Bausch, C.M. Hadi, S.H. Khan, and J.J. Lertora, "Review of the Literature and Proposed Guidelines for the Use of Oral Ribavirin as Postexposure Prophylaxis for Lassa Fever," *Clinical Infectious Diseases*, vol. 51, no. 12, 2010, pp. 1435–1441.

6. G. Stiver, "The Treatment of Influenza with Antiviral Drugs," *Canadian Medical Association Journal*, vol. 168, no. 1, 2003, pp. 49–56.

7. T.E. Ford, R.R. Colwell, J.B. Rose, S.S. Morse, D.J. Rogers, and T.L. Yates, "Using Satellite Images of Environmental

Changes to Predict Infectious Disease Outbreaks," *Emerging Infectious Diseases*, vol. 15, no. 9, 2009, pp. 1341–1346.

8. S. Altizer, R.S. Ostfeld, P.T. Johnson, S. Kutz, and C.D. Harvell, "Climate Change and Infectious Diseases: From Evidence to a Predictive Framework," *Science*, vol. 341, 2013, pp. 514–519.

9. J. Legrand, R.F. Grais, P.Y. Boelle, A.J. Valleron, and A. Flahault, "Understanding the Dynamics of Ebola Epidemics," *Epidemiology and Infection*, vol. 135, no. 4, 2007, pp. 610–621.

10. A.J. Kucharski, S. Funk, R.M. Eggo, H.P. Mallet, W.J. Edmunds, and E.J. Nilles, "Transmission Dynamics of Zika Virus in Island Populations: A Modelling Analysis of the 2013–14 French Polynesia Outbreak," *PLOS Neglected Tropical Diseases*, 2016, pp. 1–15.

11. C.E. Mills, J.M. Robins, and M. Lipsitch, "Transmissibility of 1918 Pandemic Influenza," *Nature*, vol. 432, no. 7019, 2004, pp. 904–906.

12. G. Chowell and H. Nishiura, "Transmission Dynamics and Control of Ebola Virus Disease (EVD): A Review," *BMC Medicine*, vol. 12, no. 196, 2014, pp. 1–16.

13. A. Flahault, E. Vergu, L. Coudeville, and R.F. Grais, "Strategies for Containing a Global Influenza Pandemic," *Vaccine*, vol. 24, no. 44–46, 2006, pp. 6751–6755.

14. D. Chowell, C. Castillo-Chavez, S. Krishna, X. Qiu, and K.S. Anderson, "Modelling the Effect of Early Detection of Ebola," *The Lancet Infectious Diseases*, vol. 15, no. 2, 2015, pp. 148–149.

15. O. Diekmann and J. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Chichester, U.K.: Wiley & Sons, 2000.

16. P. van den Driessche and J. Watmough, "Reproduction Numbers and Sub-threshold Endemic Equilibria for Compartmental Models of Disease Transmission," *Mathematical Biosciences*, vol. 180, no. 1–2, 2002, pp. 29–48.

17. A.S. Monto, D.M. Fleming, D. Henry, R. DeGroot, M. Makela, T. Klein, et al., "Efficacy and Safety of the Neuraminidase Inhibitor Zanamivir in the Treatment of Influenza A and B Virus Infections," *The Journal of Infectious Diseases*, vol. 180, no. 2, 1999, pp. 254–261.

18. F.G. Hayden, A.D. Osterhaus, J.J. Treanor, D.M. Fleming, F.Y. Aoki, K.G. Nicholson, et al., "Efficacy and Safety of the Neuraminidase Inhibitor Zanamivir in the Treatment of Influenzavirus Infections," *New England Journal of Medicine*, vol. 337, no. 13, 1997, pp. 874–880.

About the Authors

Mark A. Hernandez is a member of the technical staff in the Human Health and Performance Systems Group. Since joining Lincoln Laboratory in 2015, he has worked on building data pipelines and predictive models for applications in physiological signal analysis, biomedical imaging, and early detection of infectious disease. His research interests focus on developing data-driven technologies that address issues in public health. He holds a bachelor's degree in electrical engineering from the University of Illinois at Urbana-Champaign and is currently pursuing a master's degree in public health at Boston University through the Lincoln Scholars Program.

Lauren E. Milechin is a member of the technical staff in the MIT Department of Earth, Atmospheric, and Planetary Sciences. Her interests and projects involve big data, database technology, machine learning, and applications of these topics to other fields. She also assists researchers on the MIT SuperCloud supercomputing system. Previously, she worked as a technical staff member at the Lincoln Laboratory Supercomputing Center. She holds a bachelor's degree in mathematical sciences from Worcester Polytechnic Institute, where she explored mathematical modeling of disease and of population dynamics, and a master's degree in industrial mathematics from the University of Massachusetts Lowell, focusing on computer science applications, such as machine learning and algorithms.

Shakti K. Davis is a member of the senior technical staff in the Human Health and Performance Systems Group. Among her research interests is the development of sensor and signal processing algorithms that are aimed at applications for protecting warfighters against injury and enhancing warfighters' performance capabilities. Since 2015, she has supported a variety of signal processing research efforts, including characterizing exposure to hazardous military noise through the use of body-worn acoustic sensors, developing algorithms to estimate risk factors of musculoskeletal injury by using foot-mounted biomechanical sensors, and investigating classification methods that use high-resolution noninvasive physiological sensors to provide early warning of virus exposure. She joined Lincoln Laboratory in 2006 and worked on detection and classification algorithms for airborne radar systems for nine years. She holds a bachelor's degree from New Mexico State University and a doctorate from the University of Wisconsin–Madison in electrical engineering.

MARK A. HERNANDEZ, LAUREN E. MILECHIN, SHAKTI K. DAVIS, RICHARD A. DELAURA, KAJAL T. CLAYPOOL, AND ALBERT J. SWISTON

Richard A. DeLaura is a technical staff member in the Air Traffic Control Systems Group. Prior to joining Lincoln Laboratory in 2000, he was a research scientist at the University of Massachusetts Dartmouth, developing software and curriculum to introduce advanced mathematical concepts to middle and high school students. He has authored or coauthored several conference proceedings and Federal Aviation Administration research reports. He holds a bachelor's degree in chemistry and physics from Harvard University.

Kajal T. Claypool is a member of the senior staff in the Human Health and Performance Systems Group. Her current research focuses on translational informatics to understand the associations between environment, omics, and disease. She is particularly interested in using statistics and machine learning to investigate the pseudo-intervention of physical activity on obesity, aging, and Alzheimer's disease. She is also actively engaged in research and technology development in big data management and has been active in the database research community for more than 20 years. Her interests in this field include information integration, information retrieval, data analytics, and secure architectures. She has authored numerous publications in these and related areas. Prior to joining Lincoln Laboratory in 2007, she worked as a member of the technical staff at Oracle Corporation and as an assistant professor in the Department of Computer Science at the University of Massachusetts Lowell. She holds a post-doctoral master's degree in biomedical informatics from Harvard Medical School and a doctorate in computer science from Worcester Polytechnic Institute.

Albert J. Swiston was a principal investigator in the Human Health and Performance Systems Group at Lincoln Laboratory. His research focused on novel medical devices and physiological signals analysis for military and civilian applications. His work has been published in *Nature Materials*, *Science Translational Medicine*, *PLoS* publications, and IEEE journals. He served as a National Security Scholar in the U.S. intelligence community, and as a Science and Technology Policy Fellow at the National Academies' Committee on Science, Engineering, Medicine, and Public Policy. He holds bachelor's and master's degrees from Johns Hopkins University and a doctorate in materials science and engineering from MIT. He is currently the Director of Digital Operations and Innovation at Merck Research Labs.

Complex DNA Mixture Analysis

Darrell O. Ricke, Martha S. Petrovick, Catherine R. Cabrera, Eric D. Schwoebel,

and James C. Comolli

Lincoln Laboratory has developed a suite of technologies that enables rapid analysis of samples that contain DNA from multiple sources. These new techniques can process 100 million DNA sequences in five minutes. Moreover, the Laboratory's system compares a DNA profile against a database of 20 million profiles in five seconds. These capabilities are aimed at helping law enforcement and national security professionals expeditiously identify suspected criminals or terrorists, or their relatives.

Of all the leave-behind signatures for forensics- and biometrics-reliant missions, only DNA can link an individual to a crime with a high degree of certainty. Complex human DNA mixtures from three or more contributors have proven difficult to analyze with current methods. This difficulty applies especially for touch samples, with low DNA input, and the type of DNA mixtures that are often obtained in realistic criminal and intelligence investigations, such as those involving gang-owned guns, drugs, illicit cash, terrorist devices, and terrorist attack staging areas (Figure 1). Current methods are not capable of matching a reference profile to a low-concentration mixture containing DNA from four or more individuals, such as might be collected from a door knob, shared cell phone, or currency. Obtaining usable DNA profiles from a much higher proportion and much wider variety of samples than currently achievable would have a significant positive impact on forensic analyses and intelligence collection. Lincoln Laboratory is developing new tools that enable the analysis of complex DNA mixtures with up to 10 contributors. Researchers at the Laboratory are developing new technologies and analysis capabilities for unmet DNA forensics needs of the Department of Defense and law enforcement. Using a theoretical framework [1], Lincoln Laboratory has demonstrated identification of as many as 10 individuals in touch mixtures from a variety of substrates using 2,311 single nucleotide polymorphisms (SNPs), massively parallel sequencing (MPS), and advanced bioinformatics tools. With this approach, at least one person was identified at a probability of random man not excluded (P(RMNE))

DARRELL O. RICKE, MARTHA S. PETROVICK, CATHERINE R. CABRERA, ERIC D. SCHWOEBEL, AND JAMES C. COMOLLI

(a)

(b)
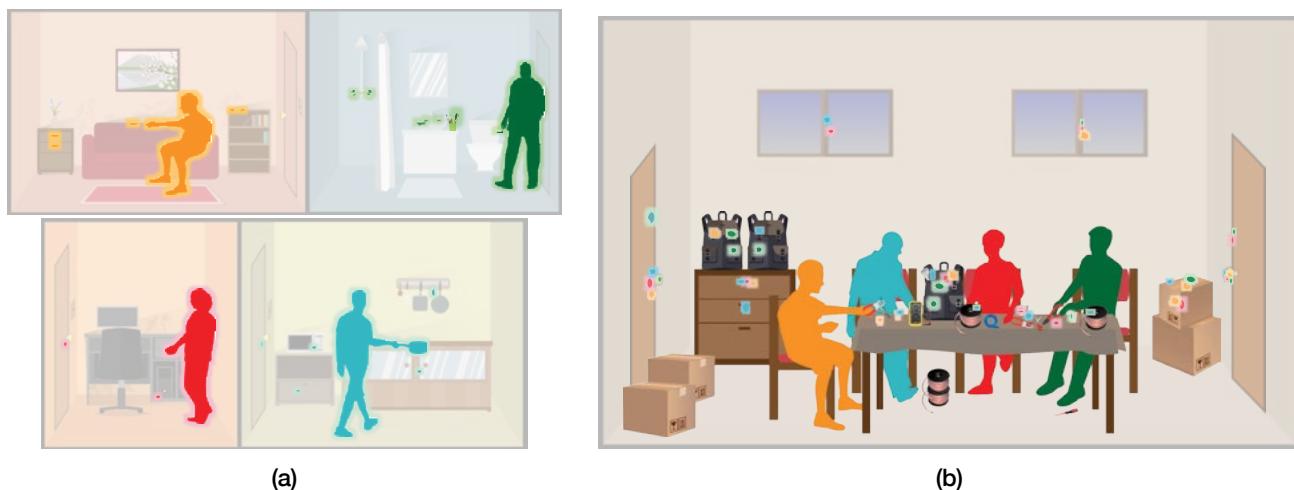
FIGURE 1. Because of limitations with the current DNA analysis methods, forensic scientists have focused their efforts on samples that are likely to be from a single source (a). However, for many realistic settings, such as a terrorist attack staging area (b), in which one needs to match a suspect's DNA, single-source samples are rare or nonexistent.
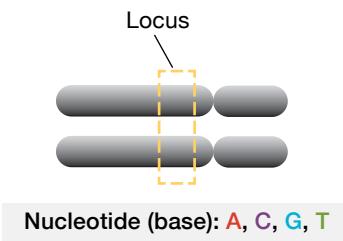
FIGURE 2. A locus is a position on a chromosome. A nucleotide, or base, is the type of monomer that is contained in the polymer that is DNA. The four types of nucleotides in DNA are commonly abbreviated A, C, G, and T.

FIGURE 3. Two different classes of DNA variants are used for forensic analysis and discussed in this article: (a) short tandem repeats (STRs) and (b) single nucleotide polymorphisms (SNPs). The STRs are repeats of four to five bases, and the number of repeats differ between people (Allele 1 and Allele 2), while SNPs vary at a single base (Allele 3 and Allele 4).

that is less than 1×10^{-9} in 97 out of 100 touch samples containing 3 to 12 contributors. Very few, if any, of these samples would be amenable to analysis using current short tandem repeat (STR) methods.

All human cells, with the exception of germline cells, contain 23 pairs of chromosomes. A specific location on a chromosome is called a locus, and a particular DNA sequence that might be found at a specific locus is called an allele (Figure 2).

Current DNA forensic techniques rely upon determining the STRs (Figure 3). An alternative approach for complex DNA mixture analysis proposed by Voskoboinik and Darvasi [1] uses a large panel of rare SNPs with low minor allele frequencies (Figure 3). On the basis of this theoretical framework, Lincoln Laboratory implemented and validated MPS SNP panels for complex DNA

FIGURE 4. Each person within the population has a unique combination of rare SNPs (shown in blue) that can be imagined as a digital barcode.

mixture analysis and kinship analysis. More specifically, we designed a panel of 2,311 SNPs with low minor allele frequencies that tend to be consistent between populations to enable complex mixture analysis across multiple ethnicities. Selection of rare SNPs (~5 percent frequency in the target population) creates unique minor allele signatures (i.e., barcodes) for individuals and enables effective differentiation of multiple barcodes in a mixture (Figure 4). Using high-throughput sequencing to generate these SNP data allows for thousands of sequence reads at each SNP locus, thereby enabling sensitive detection of minor contributors.

Experiments were performed to determine the ratio of DNA at which a contributor can be detected in two-person mixtures with this SNP panel. In contrast to the current lower limit for mixtures analyzed by STR sizing of roughly one in 20, mixture contributors have been detected down to one in 400 [2]. For standard SNP mixture analysis [3], the upper limit for the number of contributors that can be identified in the same mixture is about 9 to 10 [2]. In addition, we developed an investigative method (TranslucentID) that enables the identification of contributors in mixtures of 10 to 20 contributors [4]; see Figure 5.

After establishing that the Lincoln Laboratory MPS SNP approach to complex mixture analysis performs well on laboratory-generated, i.e., controlled, samples, an extensive set of experiments was performed to assess the method's performance with uncontrolled touch samples. Complex touch samples mixtures were generated from individuals handling different objects multiple times over the course of days or weeks. The individuals logged each

touch so that the analytical results could be compared to truth data. As shown in Figure 6, Lincoln Laboratory's MPS SNP approach successfully identified significantly more contributors to these complex mixture samples than current forensic approaches can detect, suggesting a potential paradigm shift in forensic DNA analysis that would enable successful analysis and generation of actionable data from samples that would currently go uncollected or unanalyzed.

Sample analysis time is another important consideration. Transitioning from STR-based panels with fewer than 20 loci to SNP panels with thousands to tens of thousands of loci represents computational scaling issues that that have been resolved by Lincoln Laboratory through the development of a suite of analysis tools, as described below. Analysis time includes sequence alignment, allele (DNA variant) calling, identification searches, and mixture analysis. Typically, MPS runs generate as many as 100 million DNA sequences. We developed GrigoraSNP (grígora is Greek for fast) for rapid alignment and allele calling of MPS SNP sequences [5]. GrigoraSNP typically processes 100 million sequences in approximately five minutes using four computational threads [5].

We developed FastID for rapid analysis of large SNP panels for both identification and mixture analysis [6]. FastID bit encodes SNP profiles and implements mixture analysis in three hardware instructions (exclusive OR/XOR, logical AND, and population count). FastID can compare a profile against a database of 20 million profiles in five seconds by using only a single computational thread. In addition to FastID for SNP

DARRELL O. RICKE, MARTHA S. PETROVICK, CATHERINE R. CABRERA, ERIC D. SCHWOEBEL, AND JAMES C. COMOLLI
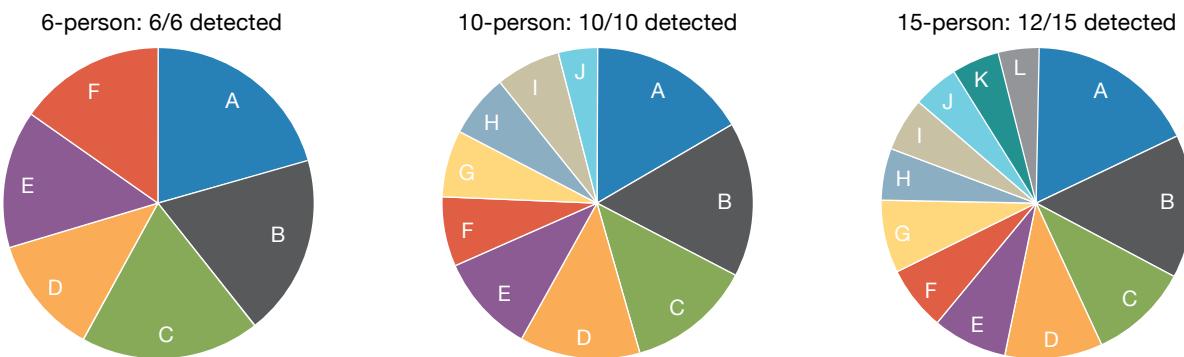
FIGURE 5. Shown are results from lab-generated equimolar mixtures of human DNA. For each pie chart, the wedge size corresponds to each subject's (identified by letter) estimated relative DNA contribution to the mixture. For the 6-person figure, the probability of random man not excluded P(RMNE) is 2.1e-54 for each of the identified individuals; for the 10-person figure, the P(RMNE) values are in the range of 5e-10 to 5e-21; and for the 15-person figure, the P(RMNE) values are in the range of 2.2e-8 to 4.5e-24.
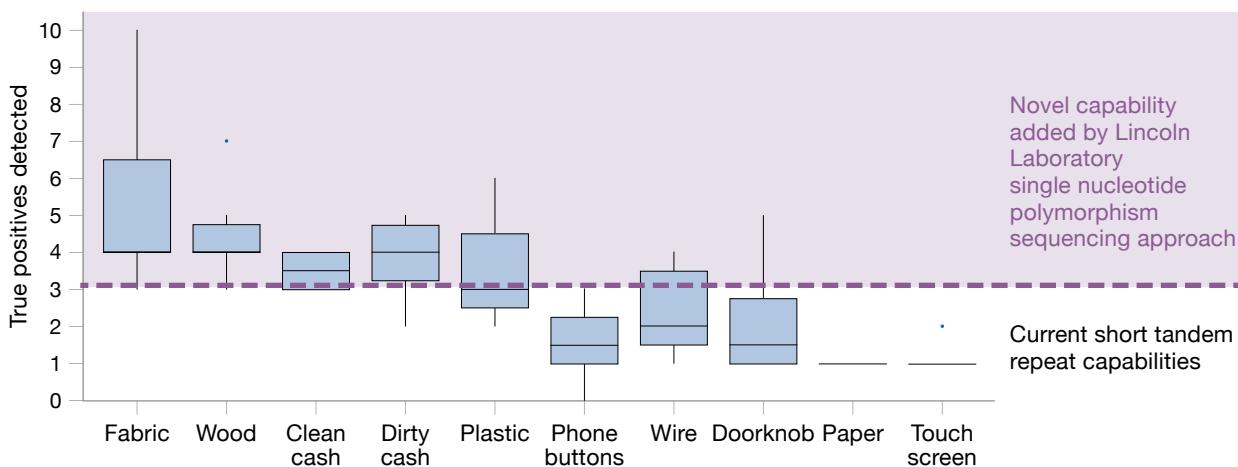
FIGURE 6. This summary of results from multiple experiments analyzing touch samples from different materials shows that the Laboratory's single nucleotide polymorphism (SNP) approach is successful with mixtures that are too complex to analyze with currently used methods.

analysis, we developed TachysSTR (tachýs is Greek for rapid) for working with STR profiles; together, FastID and TachysSTR were a 2018 R&D 100 Award winner [7] for rapid DNA forensic profile analysis. Increasing the number of DNA loci strained the statistical analysis of results in terms of computational time and numerical precision [8]. We developed Fast P(RMNE) (probability of random man not excluded) to reduce the computational time and avoid the numerical precision problems encountered with working with tens of thousands of loci [8].

In addition to developing custom methods for generating optimized SNP data and analytical tools to extract useful information from that data, Lincoln Laboratory developed the IdPrism Advanced DNA Forensics system to integrate the MPS SNP analysis tools into a graphical user interface system coupled to a relational database. As part of this system, a Linux program script runs every five minutes checking for completed sequencing runs. The binary format MPS sequence data files are transferred from the sequencer to the analysis system, converted to text format, processed by GrigoraSNPs, and uploaded to

the IdPrism database. New references are compared to all references and also compared to all mixture profiles with FastID [6] and Fast P(RMNE) [8]. New mixtures are compared to profiles of all known references with FastID and Fast P(RMNE). In addition, close relatives are detected by kinship analysis [9].

Technology advancements are enabling new DNA forensics capabilities. Our SNP mixture panel enables the characterization of complex DNA mixtures of 10 or more individuals. IdPrism has been successfully transitioned to the FBI Research Laboratory for evaluation, and discussions for commercialization have been initiated. Future expansion capabilities include customized MPS panels of tens of thousands of loci for improved prediction biogeographic ancestries for individuals, extensions of kinship for more distant relatives, and more. ■

References

1. L. Voskoboinik and A. Darvasi, "Forensic Identification of an Individual in Complex DNA Mixtures," *Forensic Science International: Genetics*, vol. 5, no. 5, 2011, pp. 428–435.

2. D. Ricke, P. Fremont-Smith, J. Watkins, T. Boettcher, and E. Schwoebel, "Estimating Individual Contributions to Complex DNA SNP Mixtures," *Journal of Forensic Sciences*, vol. 64, no. 5, 2019, pp. 1468–1474.

3. J. Isaacson, E. Schwoebel, A. Shcherbina, D. Ricke, J. Harper, M. Petrovick, et al., "Robust Detection of Individual Forensic Profiles in DNA Mixtures," *Forensic Science International: Genetics*, vol. 14, 2015, pp. 31–37.

4. D.O. Ricke, J. Watkins, P. Fremont-Smith, M.S. Petrovick, T. Boettcher, and E. Schwoebel, "TranslucentID: Analysis of Complex DNA SNP Mixtures with Large Numbers of Donors," *Australian Journal of Forensic Sciences*, published Dec. 2019 online at doi: 10.1080/00450618.2019.1699958.

5. D.O. Ricke, A. Shcherbina, A. Michaleas, and P. Fremont-Smith, "GrigoraSNPs: Optimized Analysis of SNPs for DNA Forensics," *Journal of Forensic Sciences*, vol. 63, no. 6, 2018, pp. 1841–1845.

6. D.O. Ricke, "FastID: Extremely Fast Forensic DNA Comparisons," paper in *Proceedings of the 2017 IEEE High Performance Extreme Computing Conference*, 2017.

7. L. French, "Creating 'DNA Barcodes,' Researchers Work to Speed up Forensic Analysis," *R&D Magazine* online at https://www.rdmag.com/news/2019/02/creating-dna-barcodes-researchers-work-speed-forensic-analysis, posted 14 Feb. 2019.

8. D. Ricke and S. Schwartz, "Fast P(RMNE): Fast Forensic DNA Probability of Random Man Not Excluded Calculation" [version 1; referees: awaiting peer review], 2017, available as DOI:10.12688/f1000research.13349.1.

9. B.S. Helfer, P. Fremont-Smith, and D.O. Ricke, "The Genetic Chain Rule for Probabilistic Kinship Estimation," bioRxiv. 2017, available at https://www.biorxiv.org/content/10.1101/202879v3.

DARRELL O. RICKE, MARTHA S. PETROVICK, CATHERINE R. CABRERA, ERIC D. SCHWOEBEL, AND JAMES C. COMOLLI

Appendix

# Case Study: DNA Sequencing for Forensic Geographic Attribution

Lincoln Laboratory is developing tools that use next-generation DNA sequencing technology to enhance geographic source attribution, that is, the determination of the origin of an item of interest. Knowledge gained from tracking objects back to their source, such as the site of device manufacture or assembly or the launch location of an airborne vehicle, could help characterize and eventually disrupt asymmetric threat networks. Geographic source attribution is most often used in criminal forensics for evidence and investigation, and accomplished through detailed analysis of dust. The particles within dust are derived from environmental sources, such as soil, plants, and animals, or from human activity, which are all particular to a geographic location, so can serve as a rich source of information about the site of origin. However, forensic dust analysis is currently performed by manual microscopy, so is laborious, requires extensive expertise, and takes weeks to complete. This relatively long time-to-answer greatly limits the use of geographic attribution for intelligence or tactical decision-making purposes for which an answer is needed quickly. We sought to address this limitation by developing tools that can rapidly provide attribution information from a large number of samples.

Our team focused on using a technique called metabarcoding to analyze DNA in dust on objects, in particular DNA from traces of plants and fungi. DNA barcodes are small regions of the genome that are shared by groups of organisms (in this case plants or fungi) but whose sequence differs by species. Metabarcoding involves the simultaneous analysis of multiple barcodes. By amplifying multiple plant and fungal barcodes from trace DNA extracted from dust on objects (metabarcoding) and then subjecting the barcodes to next-generation sequencing, we can rapidly obtain the signatures of plant/fungal DNA, corresponding to hundreds of species per sample, in numerous samples in parallel. These signatures were used for attribution in two ways: first for an unguided query to address the question, "Where in the world did this object reside?" and second for a guided query to compare samples from different objects to determine their common origin, in other words to address the question, "Did this sample come from a similar location as that sample?"

For unguided attribution, plant sequencing reads derived from dust metabarcoding were used to identify the originating plant species by matching to sequence records in the National Center for Biotechnology Information sequence reference database. Then, the known geographic distributions of the identified plant species were retrieved from a publicly available biogeography database Biodiversity Information Serving Our Nation (BISON). We overlaid the individual distribution maps in a geographic information system to produce a point-to-grid heat map indicating the regions where the distributions overlapped (Figure A, top). This map gave an estimate of the geographic origin of the dust sample. By comparing the estimate of our dust sample's origin to the known origin of a control dust sample, we were then able to assess the accuracy and resolution of our method.

Roughly one-third of dust samples collected from four different U.S. locations over a year provided accurate regional attribution, defined as accurate geolocation to within 600 kilometers of the site of origin. This was the first demonstration that an estimate of geographic attribution could be achieved rapidly from crude plant DNA found in dust. The accuracy of the technique depended on the number of species identified, which was significantly lower in the winter months, and on the location from which the dust sample was collected. For instance, with dust collected during spring and summer months at sites in New Mexico or South Carolina, greater than 70 percent of the samples provided accurate geolocation information. The accuracy and applicability were limited by the plant species available in both the sequence and biodistribution databases, so expansion of these databases to include more species at better resolution could result in significant improvements to the percentage of samples
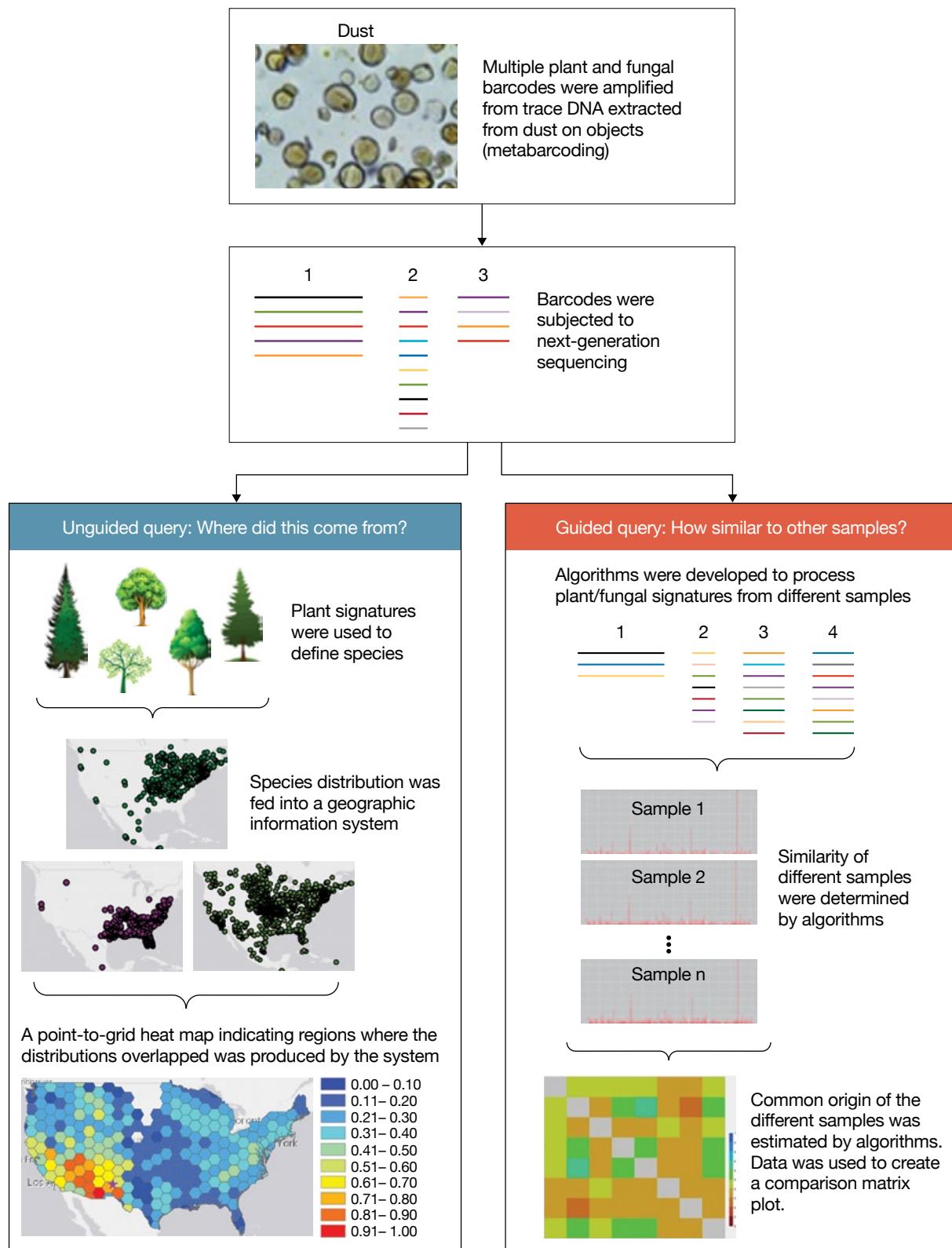
FIGURE A. Shown is a geographic attribution pipeline for an unguided and a guided query from dust samples. For the unguided query (left), plant signatures were used to define species, the distribution of which was fed into a geographic information system to produce a point-to-grid heat map indicating the regions where the distributions overlapped. For the guided query (right), algorithms were developed to process plant/fungal signatures from different samples, determine their similarity, and estimate their common origin.

DARRELL O. RICKE, MARTHA S. PETROVICK, CATHERINE R. CABRERA, ERIC D. SCHWOEBEL, AND JAMES C. COMOLLI

that yield accurate attribution and to the geolocation resolution that could be achieved.

The comparative guided attribution tool (Figure A, bottom) was designed to rapidly compare the plant and fungal sequence reads obtained from metabarcoding in different samples, determine their similarity, and estimate their common origin. Because each sample produced hundreds to thousands of different plant and fungal barcode reads, we compressed these reads into a metabarcode "signature" by employing a shingling method similar to that used for email spam detection that computationally compares signatures to determine their similarity. To test the method, we compared signatures from dust on samples that were left in Massachusetts for one week then relocated to South Carolina for one week to those from dust on samples that remained in a single location (Massachusetts or South Carolina) for one week. As expected, dust samples from objects at only one site had a significantly higher similarity to reference samples from that site than did dust samples that had been in different locations. However, samples that were in Massachusetts for as little as one day then relocated to South Carolina had similarity to reference samples from both sites, indicating that the metabarcode signature from Massachusetts was detectable after relocation. This finding suggested that the plant and fungal DNA within dust from a relocated sample enabled attribution back to its site of origin.

The two novel geographic attribution pipelines developed at Lincoln Laboratory provide initial proof of concept that environmental DNA contained within dust has utility in tracking samples and estimating their point(s) of origin. Both our work and recent technological advances in next-generation sequencing technology lead us to conclude that metabarcoding has the potential to serve as the basis for an automated system that can rapidly gather geolocation information on numerous samples or objects to provide information in a field-forward, operational setting. Further enhancement of the metabarcoding methodology, along with the expansion of the scope and accuracy of available sequence and biodistribution information, could improve attribution resolution to 50 kilometers or less within a well-referenced geographic area.

About the Authors

Darrell O. Ricke is a member of the technical staff in the Biological and Chemical Technologies Group at Lincoln Laboratory. He has broad experience in programming and software engineering, bioinformatics, molecular biology, genomics, and functional genomics. He has worked extensively on understanding the mechanisms of human diseases, biology discovery, data integration, and data mining. His current biomedical research focuses on applications in metagenomics for identification of infectious disease organisms, advanced DNA forensics, epigenetics, transcriptomics, proteomics, and disease mutation analysis. He is applying high-performance technologies to projects that include the application of current sequencing technologies to forensics and complex DNA mixtures, metagenomics, transcriptomics, epigenetics, development of multiple bioinformatics tools, and integrated data architecture design. He has been granted 10 patents and is the recipient of a 2018 R&D 100 Award for innovative methods for comparing DNA samples. He holds bachelor's degrees in computer science and in genetics and cell biology, and a master's degree in computer science from the University of Minnesota, and a doctorate in molecular biology from the Mayo Graduate School.

Martha S. Petrovick, a member of the technical staff in the Biological and Chemical Technologies Group, joined Lincoln Laboratory in 1998 to develop a cell-based sensor for biological agents and became the cell engineering leader. She also led the forensic signature science task for the Accelerated Nuclear DNA Equipment program, which was sponsored by the U.S. Departments of Defense, Justice, and Homeland Security to develop automated, rapid human DNA profiling capabilities for field biometrics and forensics applications. She currently works on improving methods for analyzing forensic DNA samples and is investigating novel markers for physical and cognitive resilience. She earned bachelor's and master's degrees in pathobiology at the University of Connecticut, and a doctorate in cell and developmental biology at Harvard University.

Catherine R. Cabrera is the leader of the Biological and Chemical Technologies Group at Lincoln Laboratory. She joined the Laboratory in 2002, initially working on hardware and software development for the identification of biowarfare agents. She was part of the team that received an R&D 100 Award for the development of the PANTHER automated cell-based bioaerosol sensor, which has since transitioned to operational use for building protection and plant pathogen detection. She currently

oversees a diverse portfolio of programs that include ones on molecular biomarkers for heath and performance, advanced DNA forensics, and engineered and synthetic biology. Her areas of technical expertise include microfluidics; biodefense technologies, systems, and architectures; red/blue team analysis; microbiome and human health; point-of-need diagnostics; genetic and epigenetic biomarkers of health and activity; and use of physiological status indicators to provide early warning of exposure to chemical warfare agents or pathogens. She holds bachelor's degrees in biochemistry and chemical engineering from Rice University and a doctorate in bioengineering from the University of Washington. Her doctoral research was focused on developing fieldable technologies to detect pathogens in resource-limited environments.

Eric D. Schwoebel is a member of the technical staff in the Biological and Chemical Technologies Group, where he has recently worked on analysis of DNA mixtures, assessment of kinship between individuals, and prediction of physical characteristics from DNA markers. He joined Lincoln Laboratory in 2001 to assist in the development of cell-based sensors for the detection and identification of biological warfare agents. He received a doctorate in cell biology from Baylor College of Medicine, worked in vaccine development at the Institute of Primate Research in Nairobi, Kenya, and returned to Baylor College of Medicine to examine the biochemistry of protein transport between the cytoplasm and nucleus.

James C. Comolli, a member of the technical staff in the Biological and Chemical Technologies Group, joined Lincoln Laboratory in 2015. His research interests include infection therapeutics and diagnostic technologies, biosensors, bacterial communities, pathogen-host interactions, and synthetic biology. Prior to joining the Laboratory, he was the leader of the Biomedical Engineering Group at Draper Laboratory; research director at ECI Biotech, a startup developing microbial diagnostics; and a staff researcher at Johnson & Johnson, working on medical devices. He has a bachelor's degree in biology from Johns Hopkins University and a doctoral degree in cellular and molecular biology from Harvard University. He completed postdoctoral training in bacterial pathogenesis at the University of California, San Francisco, and in microbial physiology at the University of Wisconsin–Madison.

Synthetic Biology

Peter A. Carr, Johanna Bobrow, James C. Comolli, Nicholas J. Guido, Frances E. Nargi,

Todd A. Thorsen, David I. Walsh, Matthew E. Walsh, Scott T. Wick, and Catherine R. Cabrera

Synthetic biologists have begun ushering in a new era of genetic engineering. In this era, how we design, build, and test new living systems is being transformed. Potential benefits span many areas, including safeguarding human health, creating new ways to make the materials we need, and responding to the diverse challenges of climate change. Synthetic biology research at Lincoln Laboratory advances the foundational technologies to enable such applications in support of national security, while preparing for potential misuse—intentional or accidental—in a world in which capabilities are advancing faster than policy and regulation.

Across all of nature, the vast number of things that living organisms can accomplish is staggering. Every living cell has the capacity to sense its environment and its own internal workings. Cells act on this information in a variety of ways, including moving, producing chemicals, signaling, reproducing, or even dying. And the instructions for these actions are stored in their DNA. Often, we can identify from an organism's genome an exact region of DNA that encodes a specific function. Genetic engineers find new ways to put those DNA instructions to work and invent new functions as well.

Dramatic advances in genetic engineering through the late 20th century often focused on a single DNA-encoded function (i.e., a gene). Many efforts focused on better understanding how living things work, such as how genes are turned on and off. Some research repurposed genes for medical applications, such as the production of human insulin by bacterial factories [1]. The newer term synthetic biology describes a shift toward applying more advanced engineering principles to how we design, build, and test living systems. These approaches have enabled researchers to use pieces of DNA from several different organisms to engineer complex systems composed of many distinct genetic functions. As synthetic gene networks were being created circa 2000, early publications drew an analogy between switch-like logical functions in molecular biology and the structure of logic circuits in electrical engineering [2, 3]. Thus, these engineered DNA-encoded systems were described as genetic circuits.

Since that time, there has been much debate about the exact definition of synthetic biology (sometimes simply

synbio) and the activities that are considered in or out of bounds. Regardless of semantics, synthetic biologists seek to advance the best practices for engineering biological systems. In doing so, they often adapt lessons learned from other engineering disciplines, such as electrical and chemical engineering. Thus, the past 20 years of synthetic biology have included grand experiments in

- Exploring standards for DNA design composition and construction
- Employing abstraction hierarchies for modular design across scales of complexity
- Developing new computer-aided design (CAD) tools for designing with DNA
- Modeling complex biological networks (circuitry) to improve and troubleshoot designs
- Characterizing failure modes (which are plentiful)
- Creating increasingly intricate and complex genetic systems
- Expanding the reach of what can be constructed (size and complexity of DNA, including entire genomes)
- Integrating with advanced capabilities, such as DNA sequencing, genome editing, miniaturization, and automation
- Transitioning these efforts from academia to industry

Synthetic biologists use these approaches to push the limits of not only how we design but also what we can design, creating new applications for engineered biological systems. Examples include programming microbes to produce medicines that are found in nature [4, 5] or to generate entirely new therapies [6]. A major area of synthetic biology progress is seen in programming simple cells (like bacteria and yeast) to biosynthesize useful molecules. These products include not only medicines but also flavors, scents, fuels, and materials. The recently released roadmap (Figure 1) from the Engineering Biology Research Consortium describes how synthetic biology impacts several economic sectors, e.g., health and medicine, energy, industrial biotechnology, environmental biotechnology, and food and agriculture [7]. These broad biotechnology-based contributions to the economy are often referred to together as the bioeconomy.

Synthetic biology research and development at Lincoln Laboratory supports national security in three ways. First, we engineer living systems that meet a specific need, such as responding to a biothreat like anthrax. For
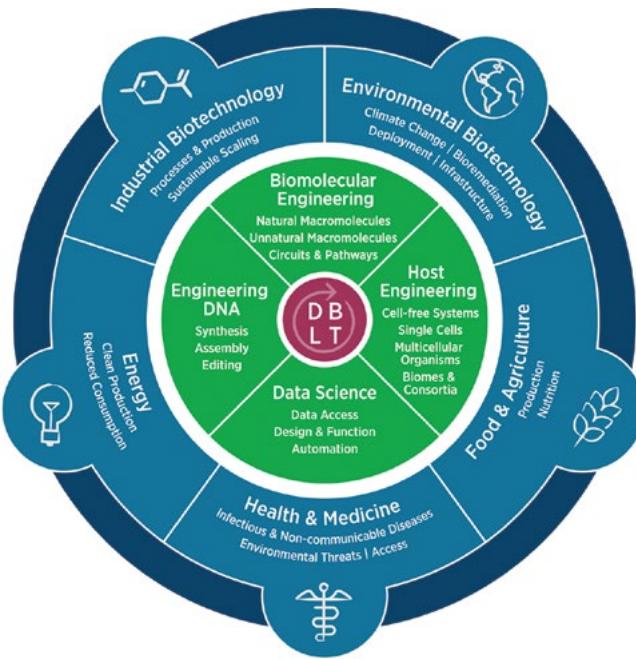
FIGURE 1. The Engineering Biology Research Consortium roadmap highlights not only key economic sectors that contribute to the overall bioeconomy but also their technical underpinnings.

example, the CANARY sensor system employs a genetic circuit that produces light within seconds of coming into contact with a biothreat, giving detection that is both rapid and specific. The engineered cell lines at the heart of CANARY can be flexibly reprogrammed to respond to different pathogens, including new or emerging diseases.

Second, as synthetic biology is still a young discipline, we create foundational tools that support our own applications and the field as a whole. Many of our efforts focus on prototyping, such as engineering better ways to assemble the large pieces of DNA that encode our genetic designs, and measuring the effectiveness of these designs. Our projects in microfluidic DNA assembly have demonstrated how the biochemical reactions used for genetic circuit construction can be miniaturized to volumes one-thousandth of that typically used in a standard lab. Our PERSIA system provides a means to measure RNA and protein production in real time, employing cell-free reactions for prototyping that avoid otherwise laborious experiments in live cells. Contributing to the foundations of synthetic biology includes fostering norms of responsible research. Our contributions also include broadly supporting the

PETER A. CARR, JOHANNA BOBROW, JAMES C. COMOLLI, NICHOLAS J. GUIDO, FRANCES E. NARGI,
TODD A. THORSEN, DAVID I. WALSH, MATTHEW E. WALSH, SCOTT T. WICK, AND CATHERINE R. CABRERA

development of synbio hardware through open-source design sharing and enabling the use of shared fabrication resources, such as in makerspaces.

Third, we analyze the technology landscape of synthetic biology and its impact on the world around us. We do this to better advise the U.S. government on synthetic biology issues and to prepare ourselves for future opportunities and challenges. These efforts often consider how synthetic biology will impact biodefense, human health and performance, or manufacturing. Such considerations apply as well to the overall growing bioeconomy, and ways to support and protect it. Some analyses focus on immediately available technology for meeting current needs while others anticipate the evolution of the technologies and their impacts several decades into the future.

Programming Biology

The DNA of a living cell acts as the instruction set for a large collection of highly connected control elements (those that switch genes on and off) and actuators (such as enzymes that perform catalysis). At a high level of abstraction, many scientists have noted similarities between these DNA-based instructions and other types of engineered systems, such as computer code and electronic circuits (digital and analog). Thus, many synthetic biologists model their DNA-based systems as genetic circuits to better understand and engineer them.

To achieve the representation of genetic systems as genetic circuits, synthetic biologists make use of ideas developed in computer science, information theory, and electronic circuit design. The commonality is that both electronic systems and living systems can achieve a multitude of possible states in which they can exist. Electronic circuits deal primarily with controlling the states and fluxes of electrons. Genetic circuits deal primarily with states and fluxes of molecules. Both can be represented in terms of digital and analog logic. With digital electronic logic, you can represent discrete states with a switch that is open or closed, or a capacitor that is charged or uncharged, or a transistor that allows current to flow or not. In a genetic system, the state can be if a gene is turned on (producing RNA and often translating a protein), if a protein is bound to a specific DNA molecule, or if a molecule in the cell has been chemically modified. But it is important to note that representing

biological systems in digital states can sometimes be a problematic simplification.

Along with the experimental capability to engineer biological systems came some of the first genetic circuit designs that were created to implement logic circuits [2, 3]. Figure 2 shows different representations of a logical OR function. If either input is present, then the OR function returns a specific positive result (i.e., on, true, or 1). The schematic for the genetic OR circuit indicates components encoded in DNA, which produce their output in the presence of chemical inputs. If either chemical input is present, the observable result is that the system produces light (bioluminescence). Importantly, whether the OR is represented in computer code, electrical switches, or genetic logic, several different designs can be used to achieve the desired OR behavior. For an electrical circuit, this behavior might be achieved with switches (shown here), diodes, or transistors. For a genetic circuit, one might use gene activators (proteins that turn on gene expression, as shown), repressors (proteins that turn off gene expression), or protein modification events. Which of these choices proves most effective for a given design goal will depend heavily on specifications and context, such as the environmental conditions in which the engineered system will function, and desirable traits of the sensor, such as rate of response.

Figure 3 shows the genetic OR example expanded in more detail. The desired function is a cell that can detect both lead and mercury contaminants in water, emitting light if either contaminant is present (or if both are). The DNA-encoded program is typically shown arranged along a single line, representing the (linear) physical organization of the DNA parts within a large DNA molecule. Table 1 gives details of key DNA parts that make up the OR circuit. Synthetic biology approaches have the potential to make some aspects of this project much easier. For example, several useful functional units are already available in existing collections of DNA parts, such as the Registry of Standard Biological Parts [8] and Addgene, including previous users' experience with the performance of DNA parts. Design tools such as Cello can help the user explore possible device architectures and simulate their behavior before moving to the step of physically building the circuit [9].

There can also be pitfalls to representing genetic devices with such simple levels of abstraction. Over-reliance

on simplification with these designs can mask the relevance of the complex biological context in which the device does its job. For example, the components shown for the lead OR mercury detector of Figure 3 may interfere with related systems in the host cell (often referred to as the chassis) that also bind to and remove toxic metals. Many such designs require a significant amount of tuning or redesign to take the cellular context into account.

(a)

(b)

(c)
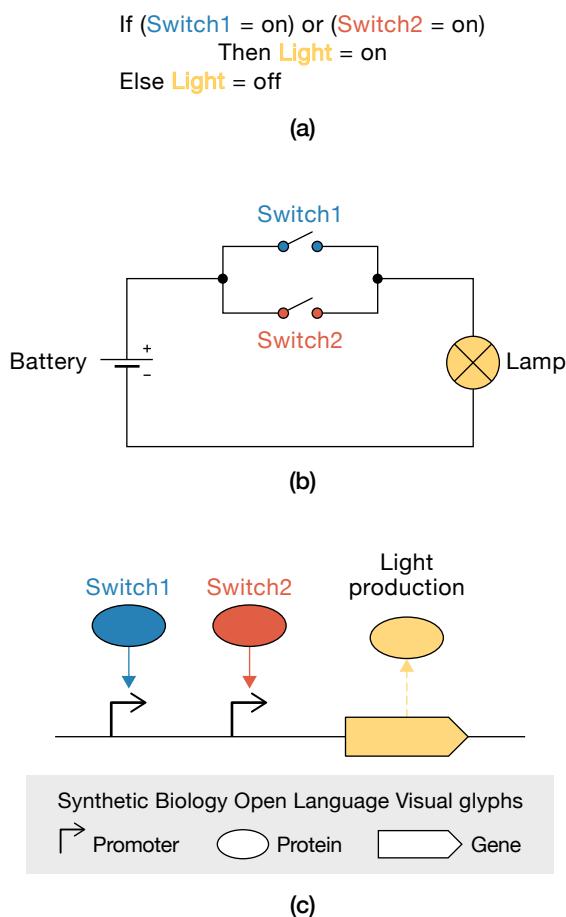
FIGURE 2. Different versions of an OR function are shown. In each case, if either switch 1 (blue) or switch 2 (red)—or both—are on, then the output (light) is turned on. With computer code (a), this action can be as simple as employing the term "or" in a logical statement. The electrical OR shown (b) is made with switches, that when closed allow current to flow. The genetic OR circuit (c) is made from DNA-encoded proteins that activate DNA control elements, called promoters, to express genes. Some of the DNA components of the design are shown arranged along a straight line including promoters and one gene. This simplified genetic OR uses two different proteins (switch 1 and switch 2) to detect dangerous chemicals, turning on production of another protein that generates light (bioluminescence).

Microfluidics

Microfluidics are devices that manipulate very small amounts of fluid—microliters, nanoliters, and even picoliters. (For reference, a single drop of water is about 50 microliters.) Often referred to as lab-on-a-chip technology, microfluidics can be used to integrate several different operations that would otherwise be performed individually by hand in the lab. The small volumes involved are an advantage for requiring only a minimum of reagents that are often costly or difficult to acquire. When handling hazardous materials, using only a microfluidic volume can also enhance safety by drastically reducing the total amount of a dangerous chemical or biological agent required. The small feature size of a microfluidic device

(a)

(b)

(c)
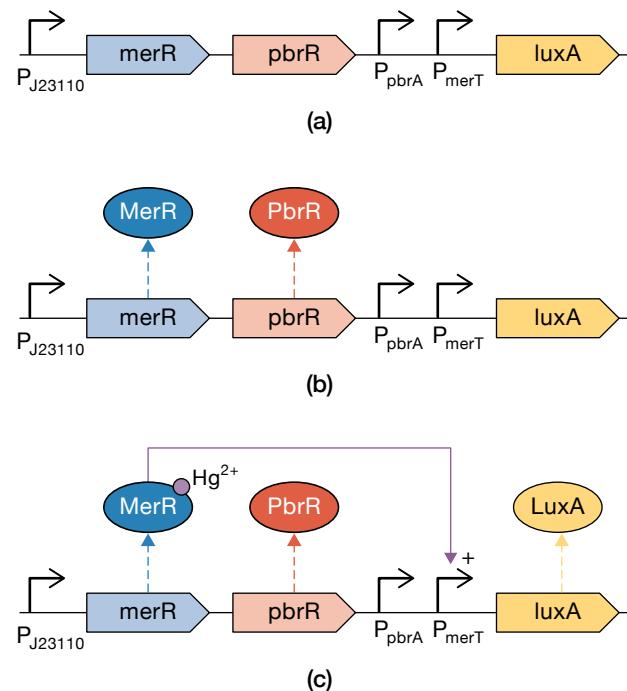
FIGURE 3. Illustrated is a simplified genetic OR circuit to detect lead and/or mercury (a). The layout of the DNA components represents the physical molecule of DNA that encodes these parts. These include three control elements (promoters) and three protein-encoding genes. These DNA parts are described in more detail in Table 1. Once the DNA instructions are inserted into a cell, the DNA-encoded regulator proteins MerR and PbrR are produced at a constant level because promoter P_{J23110} is always on (b). If mercury ions (Hg^{2+}) are present, they will bind to protein MerR, activating promoter P_{merT}, leading to the production of the LuxA protein (c). LuxA uses chemicals present in the cell (made by additional Lux proteins, not shown) to generate light. Detection of lead ions (Pb^{2+}) via the PbrR protein and the P_{pbrA} promoter works similarly.

PETER A. CARR, JOHANNA BOBROW, JAMES C. COMOLLI, NICHOLAS J. GUIDO, FRANCES E. NARGI,
TODD A. THORSEN, DAVID I. WALSH, MATTHEW E. WALSH, SCOTT T. WICK, AND CATHERINE R. CABRERA

also facilitates scale-up in numbers of parallel operations: in some cases, more than 10,000 microfluidic reactors can function side by side in an area of no more than a few square centimeters [10]. This compact format, combined with multiple ways to control on-chip activities, allows multiple different processes to be combined in the same device. Several Lincoln Laboratory projects marry synthetic biology with microfluidics by integrating functions such as DNA construction, installation of DNA designs into living cells, and use of DNA to perform cell-free synthesis and other enzymatic reactions.

Figure 4a shows one element of a very simple microfluidic device. Multiple layers of flexible, transparent silicone rubber (polydimethylsiloxane) have been patterned and assembled by using the techniques of multilayer soft lithography, and sealed to a rigid glass coverslip (bottom). A single narrow microchannel connects a fluid inlet and a fluid outlet. Just above this channel are two more perpendicular channels. These top channels can be pressurized with air or liquid, causing them to expand and press down on the single channel below, acting as valves to seal it off from the inlet and outlet. The sealed channel can then be used like a miniature test tube to perform a biochemical reaction, culture and observe live cells, and more. These simple design elements can be repeated and recombined in many ways

to integrate complex multistep processes. We employ custom hardware and software to actuate arrays of control valves, creating an elaborate fluidic trafficking system that can manage many processes in parallel or address individual reactors as needed. Figure 4b shows a Lincoln Laboratory microfluidic device that integrates 96 of these reactors. These reactors can be used flexibly, such as for assembling DNA, synthesizing protein in a cell-free expression system, or monitoring an enzymatic reaction. Many other kinds of microfluidic devices exist and use a diverse array of materials. Some microfluidics can reach a very high degree of feature density and complexity [10].

We are engineering new technology to miniaturize and integrate the processes for constructing DNA and using that DNA, specifically combining biological and biochemical processes with microfluidic hardware. A crucial example is the process of DNA construction, at the heart of most synthetic biology research and development. One of our methods utilizes an automated droplet-spotting robot to place ultra-small volumes (as low as 300 picoliters) of DNA solutions on an epoxy-coated microscope slide, followed by plasma bonding to a microfluidic channel architecture made from polydimethylsiloxane. A reaction mixture for performing DNA assembly is pushed through the reactor channels to rehydrate, mix, connect, and recover DNA from parts to a complete

Table 1. DNA Parts for the Genetic OR Circuit for Detecting Lead and Mercury

NAME	FUNCTION (MOLECULE)	DETAILS	ORGANISM SOURCE	REGISTRY PART NUMBER
pbrR	Regulator (protein)	Lead-responsive	*Cupriavidus metallidurans*	BBa_K1701001
merR	Regulator (protein)	Mercury-responsive	*Shigella flexneri*	BBa_K1420004
P_{J23110}	Promoter (DNA)	Medium-strength constitutive (always on) promoter	Anderson promoter collection, derived from *Escherichia coli*	BBa_J23110
P_{pbrA}	Promoter (DNA)	Regulated by PbrR protein and Pb^{2+}	*Cupriavidus metallidurans*	BBa_I721001
P_{merT}	Promoter (DNA)	Regulated by MerR protein and Hg^{2+}	*Shigella flexneri*	BBa_K1758342
luxA	Enzyme reporter (protein)	Luciferase (light-generating) protein	*Vibrio fischeri*	BBa_K785003

assembly. Throughout this work, we have developed separate hardware (programmable pneumatic pumps) for controlling microfluidics, an alignment system to accurately bond device layers, and valve control strategies for device filling, reaction isolation, mixing, and recovery.

Biosensors: CANARY

Nature has produced many ways to sense the world, including vision (detecting light), smell (detecting molecules), and touch (perceiving hot, cold, pressure). Living things also have ways to sense inside their own bodies—such as with immune systems that respond to unwelcome invaders. Synthetic biologists look at these examples and see opportunities to engineer new systems from the biological parts that compose the heart of these senses.

An early example of synthetic biology–style engineering at Lincoln Laboratory was the creation of the CANARY (Cellular Analysis and Notification of Antigen Risks and Yields) sensor [11, 12]. The need was for a system that could rapidly and specifically detect different pathogens or toxins, such as the high-risk biological agents

that compose the U.S. Select Agents and Toxins List [13]. The concept was to harness the exquisite combination of both variation and specificity of the immune system. Cells derived from the mouse immune system (B cells) were programmed and employed to detect different agents and respond rapidly by producing light (Figure 5). A suite of CANARY sensors has been produced to detect bacteria, viruses, toxins, and DNA molecules.

The basic CANARY system is designed to express antibodies on the surface of the B cells. These antibodies detect antigens that are expressed in multiple copies on the surface of pathogens. In the case of a bacterium, there may be 10,000 copies of a given antigen while a virus may carry about 1,000 copies. Toxins are generally monovalent (1 copy) and thus would not be detected by the basic CANARY sensor outlined here. However, a simple change to the basic design involving expression of multiple different antibodies in a single cell line has now enabled CANARY to detect these medically important toxins as well.

While the term synthetic biology had not yet been popularized, crucial elements of a synthetic biology design-build-test engineering cycle were notable in this

(a)

(b)
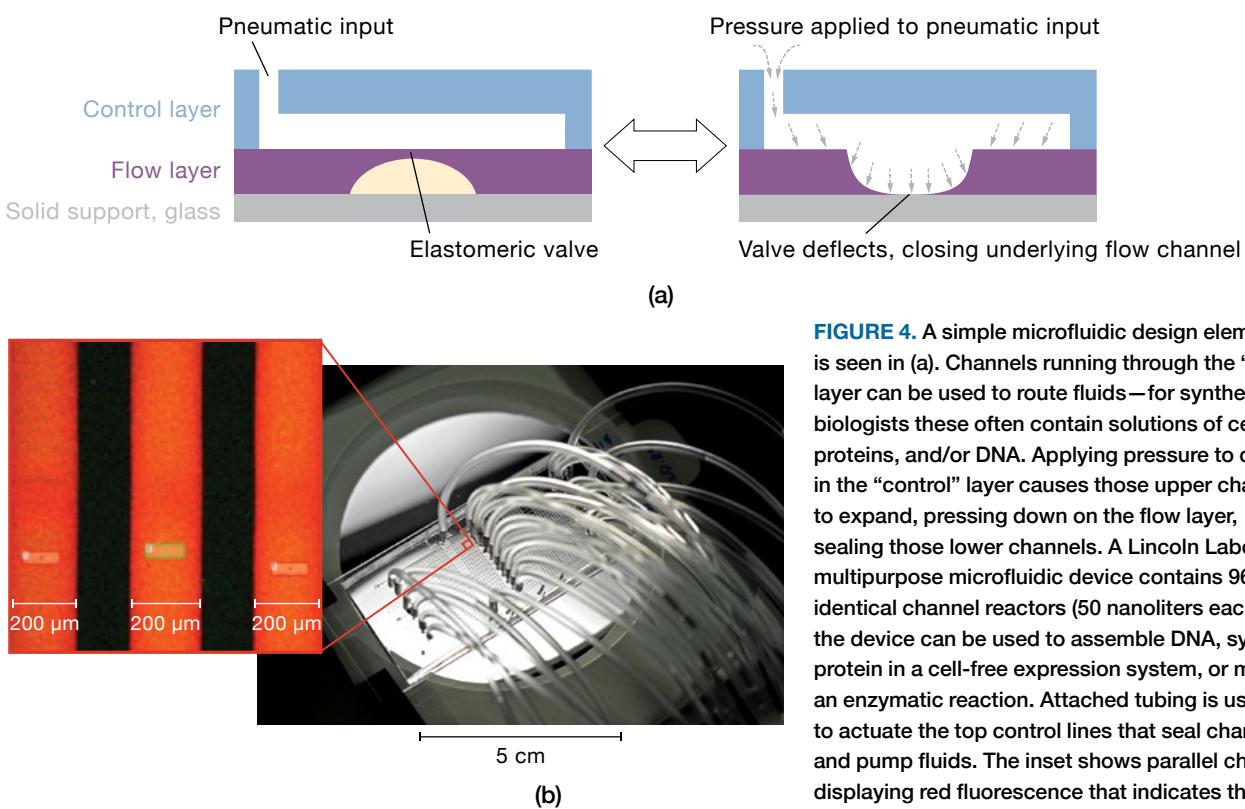
FIGURE 4. A simple microfluidic design element is seen in (a). Channels running through the "flow" layer can be used to route fluids—for synthetic biologists these often contain solutions of cells, proteins, and/or DNA. Applying pressure to channels in the "control" layer causes those upper channels to expand, pressing down on the flow layer, sealing those lower channels. A Lincoln Laboratory multipurpose microfluidic device contains 96 identical channel reactors (50 nanoliters each) (b); the device can be used to assemble DNA, synthesize protein in a cell-free expression system, or monitor an enzymatic reaction. Attached tubing is used to actuate the top control lines that seal channels and pump fluids. The inset shows parallel channels displaying red fluorescence that indicates the cell-free production of protein.

PETER A. CARR, JOHANNA BOBROW, JAMES C. COMOLLI, NICHOLAS J. GUIDO, FRANCES E. NARGI, TODD A. THORSEN, DAVID I. WALSH, MATTHEW E. WALSH, SCOTT T. WICK, AND CATHERINE R. CABRERA

project. First, the important real-world specifications were identified, namely the capability to respond quickly, the flexibility to reprogram the system to detect distinctly different threats, and the need for both sensitivity and specificity. Next, a cellular chassis (choice of organism and cell type) was selected as the runtime environment for the genetic program, taking maximum advantage of existing natural features. And finally, genetic parts from several different organisms were integrated to provide the capabilities needed to complete the system. Iterative redesign cycles were used both to improve system performance and to adapt CANARY to detect diverse threats.

The genetic programming of CANARY works much like an IF/THEN computing statement: IF the target agent is present, THEN produce light. The parts required to perform this task are shown in Figure 5, along with the process of how CANARY senses and reports. We note that all proteins needed by the CANARY system are synthesized and present before any detection event occurs, with no additional turning genes on or off required. This feature allows CANARY to employ a much faster process than that used by the circuit shown in Figure 3, enabling CANARY to respond to the sample in seconds.

Hardware to monitor the light output of the engineered CANARY cells was also developed at Lincoln Laboratory. Figure 6 shows the timeline of development, including prototypes for the TCAN (Triggered CANARY) and PANTHER (Pathogen Analyzer for Threatening Environmental Releases) detection systems. The CANARY technology and PANTHER were both transitioned to industry. Lincoln Laboratory's ability to co-develop both wetware (living cells, or biochemical reactions) and hardware allows us to pursue new innovations in synthetic biology that can reach beyond the laboratory and be integrated into real-world, fieldable devices.

Lincoln Laboratory has continued to develop CANARY technology since 1997, including modification of mouse B-cell lines to improve their durability and ability to remain viable even after two weeks stored at room temperature. Current CANARY projects support the U.S. Department of Agriculture needs to detect plant pathogens for protecting U.S. agriculture. Other ongoing
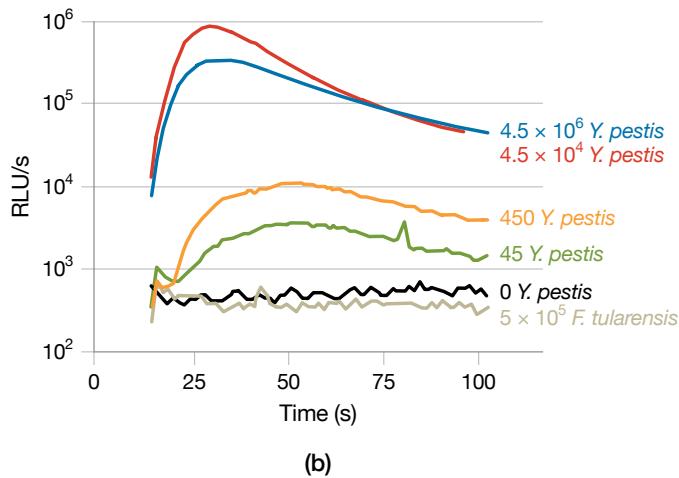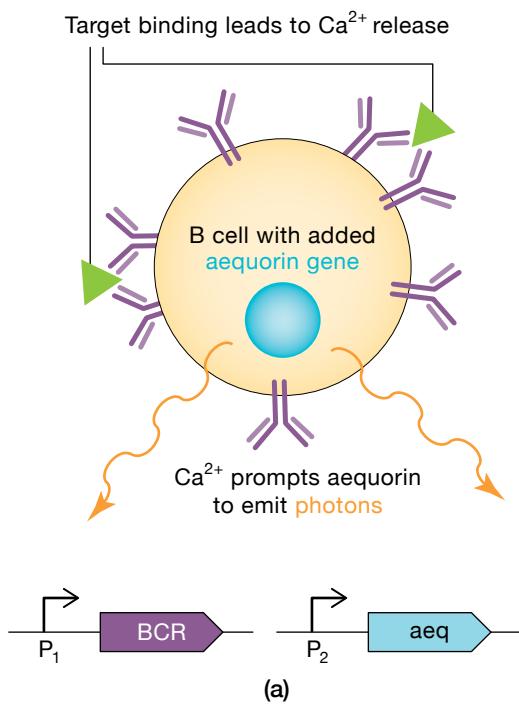
FIGURE 5. The sensing element in CANARY is a mouse B-cell line that has been genetically engineered to produce both the aequorin protein in the cytoplasm, and B-cell receptors (BCR) displaying recombinant antibodies on the cell surface (a). Aequorin, cloned from the same jellyfish that produces green fluorescent protein, is a calcium-requiring luminescent protein that emits photons. When these engineered cells are exposed to the specific bioagent they are engineered to detect, antigens on the bioagent bind to multiple BCR proteins, crosslinking them and initiating a cascade of molecular events that causes an increase in the calcium concentration inside the cell. This calcium activates the aequorin to produce light. The light can be detected by a photomultiplier tube or other hardware-based detector. Sample data are shown in (b), indicating the sensitivity and specificity of CANARY cells engineered to detect *Y. pestis* (plague) within 30 seconds of adding the sample. RLU = relative light units.
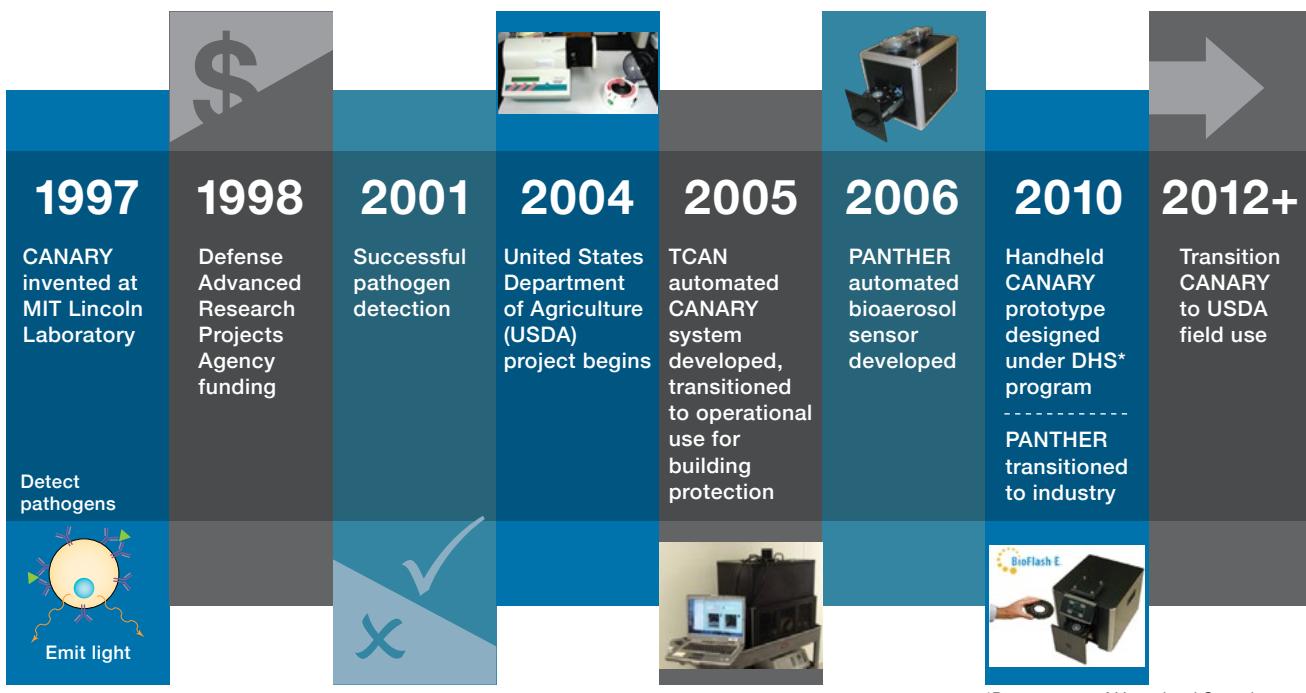
FIGURE 6. The timeline for the development of CANARY and its subsequent iteration as PANTHER spanned more than a decade; updates to the technology continue to be made.

Lincoln Laboratory synthetic biology projects develop cell-based and molecular biosensors for a variety of U.S. government needs.

Accelerating Build and Test: A Biomolecular Prototyping Unit

As with other engineering disciplines, designing a biological system typically requires an iterative process of prototyping, troubleshooting, and redesign. Thus, many synthetic biologists devote themselves to creating tools to speed up the design-build-test cycle. At Lincoln Laboratory, we are engineering an integrated pipeline that combines the capabilities for making DNA and putting it to work, namely, the Biomolecular Prototyping Unit (BPU, Figure 7).

The BPU concept represents the integration of key processes needed for the build and test phases of the synthetic biology engineering cycle. With design files in hand, a user would send a request to a BPU pipeline. For a synthetic biologist, the DNA fabricated might represent a design (or several) for a new genetically encoded sensor circuit. But the BPU will also be valuable for biologists building DNA molecules with sequences of natural origin, possibly pulled from a database, encoding a protein they wish to study, even if they have never seen the source organism in real life. For medical clinicians, the DNA sequences might come from sequencing a patient's DNA, or the DNA of a pathogen infecting that patient, and the question may be one such as, "What are the best medicines to treat this specific person's infection?"

Crucial elements of the first BPU module, DNA construction, are (1) chemically synthesizing short strands (called oligonucleotides, or simply oligos, for example, 60 DNA bases long); (2) using enzymes to assemble these short pieces into longer cassettes (for example, 1,000 DNA base pairs long, a common length scale for a single gene); and (3) dealing with manufacturing defects in the DNA molecules (i.e., quality control and assurance). Many different approaches can be used for accomplishing these steps, and some steps may be required more than once. For example, building a 5,000-base-pair piece of DNA may require assembling many oligos 60 bases long into units 1,000 base pairs long, and then joining several of those units to reach 5,000 base pairs. Quality control and assurance steps may include using a special enzyme that degrades and removes defective DNA, while another

PETER A. CARR, JOHANNA BOBROW, JAMES C. COMOLLI, NICHOLAS J. GUIDO, FRANCES E. NARGI, TODD A. THORSEN, DAVID I. WALSH, MATTHEW E. WALSH, SCOTT T. WICK, AND CATHERINE R. CABRERA
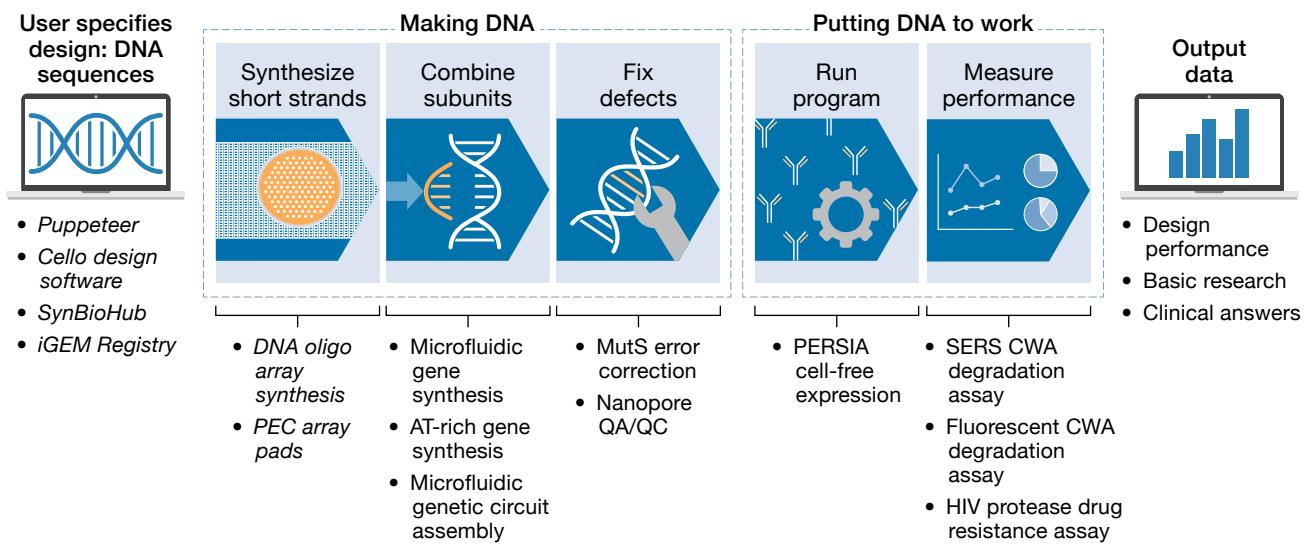
FIGURE 7. Lincoln Laboratory's vision for the Biomolecular Prototyping Unit (BPU) pipeline is illustrated here. Accepting DNA design files specified by the user, the BPU would physically construct those DNA molecules, using modules shown in the "making DNA" portion of the pipeline. Then those DNA molecules would be "put to work" either inside living cells or by using cell-free expression systems. Below these different modules are listed efforts at Lincoln Laboratory that contribute to the overall development of the BPU. (Collaborators' and related efforts are listed in the first two columns in italics.) QA/QC = quality assurance/quality control; SERS = surface-enhanced Raman spectroscopy; CWA = chemical warfare agent; PEC = photoelectrochemical.

approach can simply be sequencing enough samples to find one that is error-free.

The second BPU module, putting DNA to work, represents a large number of possible choices for how DNA-based designs are implemented and measured. An early step is typically to insert the DNA into a living cell, which then runs the program. Cells can be from bacteria, yeast, humans, or a variety or other organisms, and the design must be matched to the type of cell being used. Similarly, performance can mean many different things, such as production of a useful chemical, emitting a warning signal, or reporting on the inner workings of the cell. The tools needed to measure these functions are also quite diverse. Often a significant part of the design is to convert some cellular process into a signal that can be observed with available hardware, such as a fluorescence or bioluminescence detector, mass spectrometer, or DNA sequencer.

Our goal for an effective BPU is to maximize integration and miniaturization while reducing the total expertise required of the user. Some larger-scale, lower-throughput examples of this concept have been demonstrated very effectively, such as the Digital-to-Biological Converter produced by a team at the J. Craig Venter Institute [14]. The BPU

vision in Figure 7 includes contributions from published results and ongoing projects of Lincoln Laboratory and our network of collaborators. One example is our production of a microfluidic genetic assembler, capable of combining multiple DNA parts, using a variety of different common DNA construction approaches (Figure 8) [15]. We showed that four common methods (restriction/ligation, Golden Gate, Gateway, and Gibson Assembly) are each compatible with our microfluidic devices and can be miniaturized more than 20-fold compared to common laboratory practices. Some of these methods showed potential for further 100-fold minimization in reaction volume. This research informed our current work developing high-throughput DNA assembly devices.

Measurement: We Want a Multimeter

For the synthetic biologist, there is no one simple multipurpose measurement tool akin to an electrical multimeter (or oscilloscope). Instead, there exists a suite of powerful (often expensive) measurement tools, such as flow cytometers, DNA sequencers, quantitative polymerase chain reaction (qPCR) thermocyclers, and electrophoresis instruments. Many of these rely on converting the biological effect we want to observe into

a more easily measurable signal, such as fluorescence. In fact, all four instruments noted above rely on fluorescence for reading out answers, and all four have shown great adaptability for posing and answering many different kinds of questions. Over time, many of these tools have become faster, easier to use, and/or more affordable. But the question remains of how close we can get to the properties of a multimeter—capable of quick measurements, switchable between detection modes (volts, amps, ohms, farads), portable, and cheap.

(a) (b) (c)

(d)

(e)
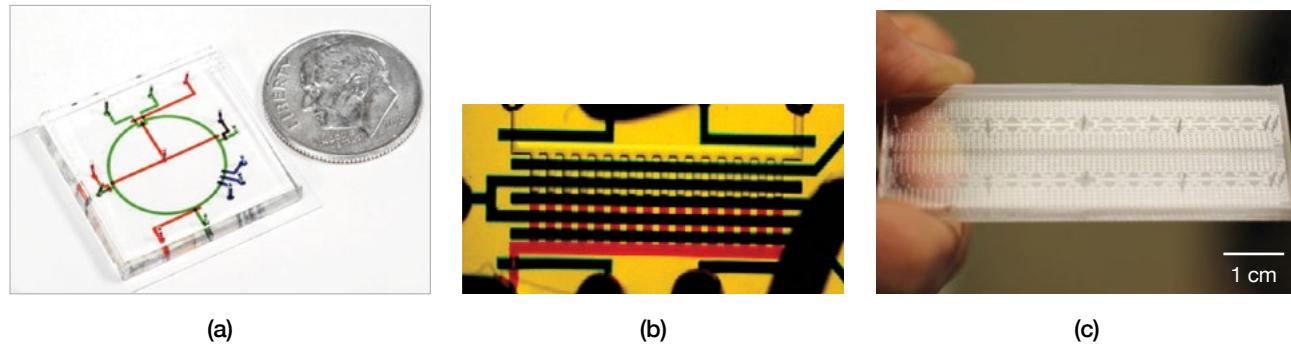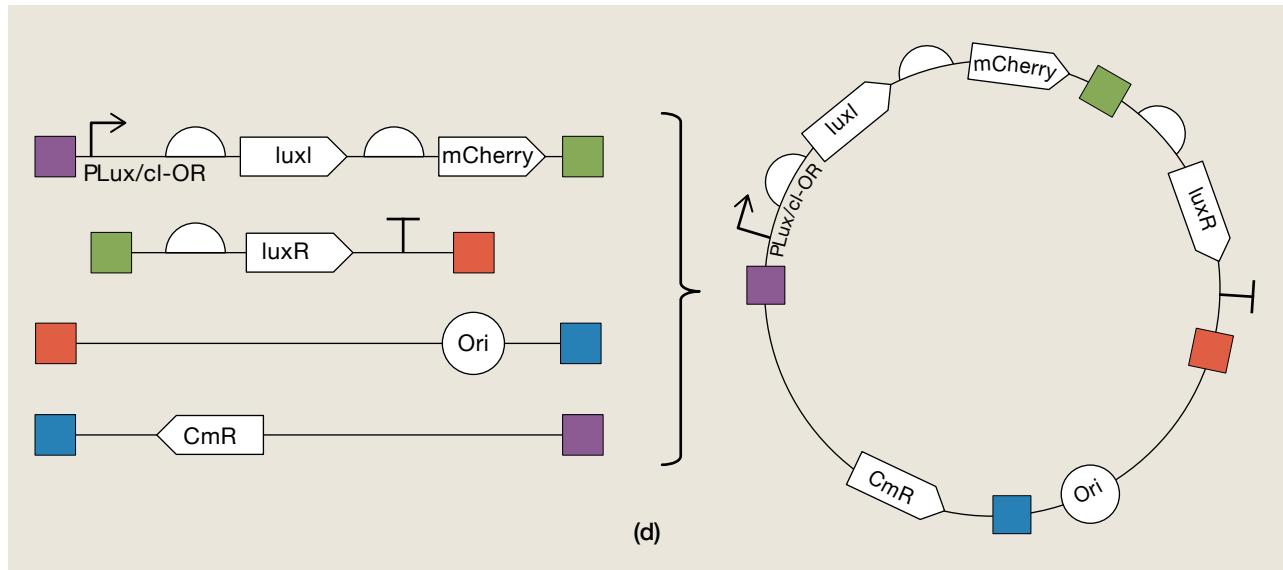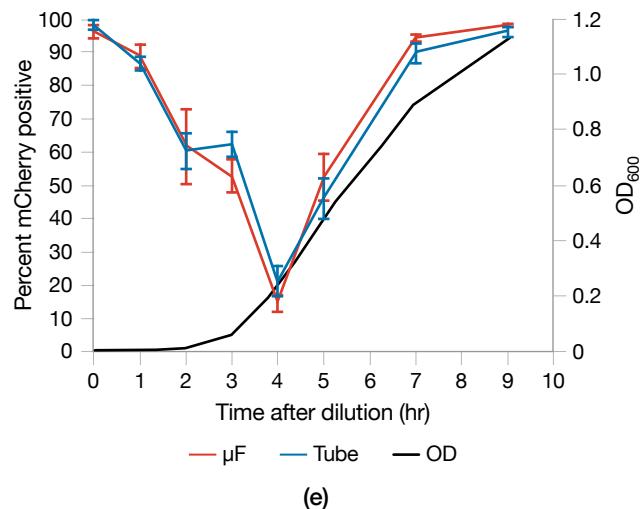
FIGURE 8. Various microfluidic genetic assemblers developed at Lincoln Laboratory are shown in (a–c), capable of performing either a single reaction (a, 1-plex), 16 parallel reactions (b), or 256 parallel reactions (c). For the microfluidic assembly of genetic circuits, several different genetic circuits were constructed using the 1-plex mixer, testing multiple assembly approaches, including the quorum-sensing circuit shown in (d) (employing the Gibson assembly method). Genetic circuits assembled in the microfluidic device (µF) performed comparably to those assembled in conventional test tube (Tube) reactions (e), in this case producing a fluorescent protein (mCherry) in response to sensing high cell density (OD) [15].

PETER A. CARR, JOHANNA BOBROW, JAMES C. COMOLLI, NICHOLAS J. GUIDO, FRANCES E. NARGI,

TODD A. THORSEN, DAVID I. WALSH, MATTHEW E. WALSH, SCOTT T. WICK, AND CATHERINE R. CABRERA

Our most recent contribution to the idea of a bio-multimeter is called PERSIA [16]. This labeling and detection scheme is named after its components, PURExpress-ReAsh-Spinach In-vitro Analysis. PERSIA provides a way to monitor the molecular events in a complex cell-free reaction while those events are occurring.

Cell-free transcription and translation reactions in general are useful for reproducing the central dogma of biology: DNA is read to make an RNA copy; the RNA copy is translated to make a protein. These reactions can be used to try out an idea, or answer a simple question, much faster than working in live cells. (However, cell-free reactions do not recreate all the complex interactions that are present in a live cell.) Increasingly, synthetic biologists are using such systems to develop sensors, prototype genetic circuits, and perform small scale point-of-need manufacturing.

As shown in Figure 9, the components needed for PERSIA are encoded into the DNA design for one or more constructs. In addition to the DNA sequence for
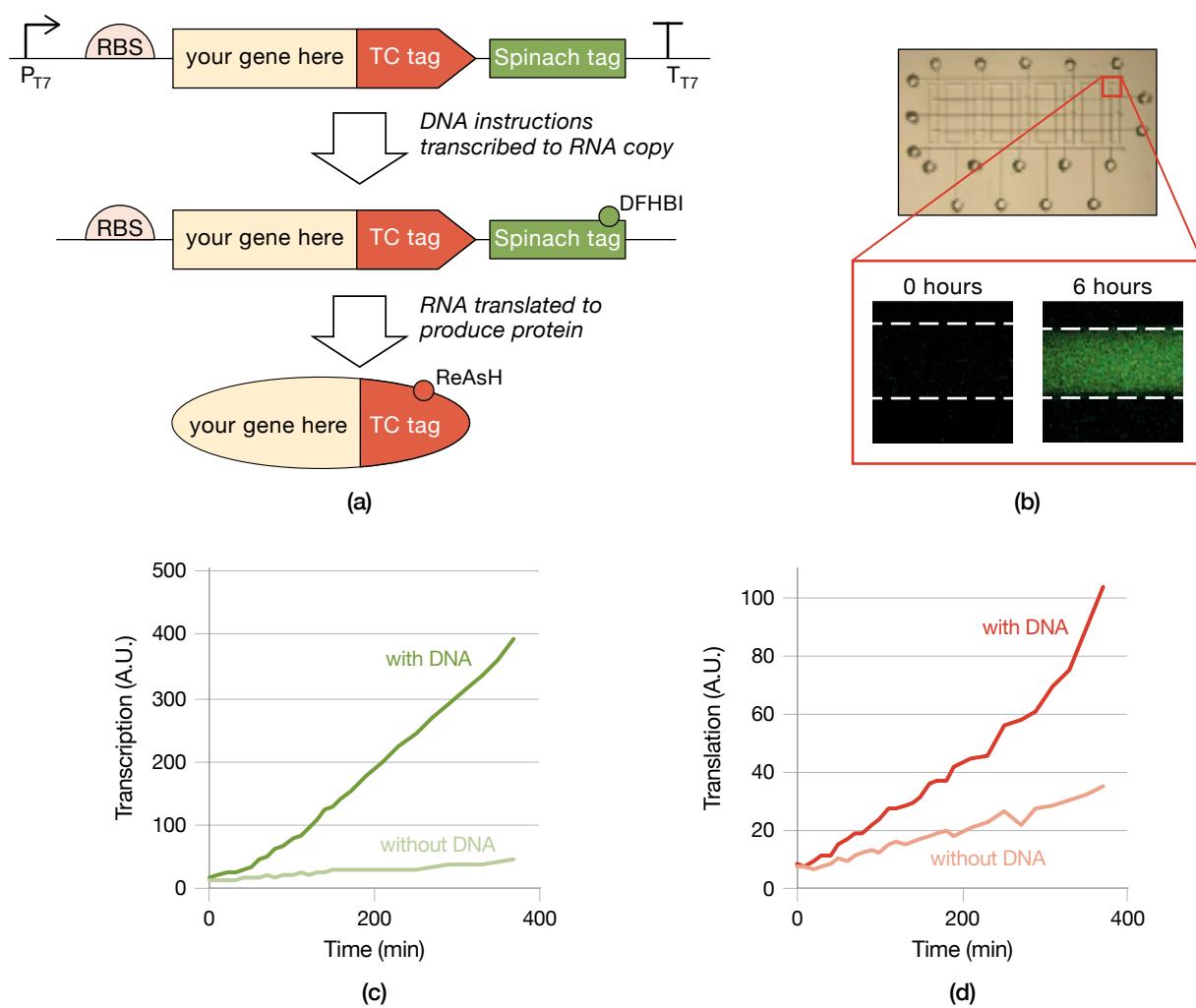
FIGURE 9. Typical design layout of DNA used with PERSIA is shown in (a). RNA polymerase start (P_{T7}) and stop (T_{T7}) signals from bacteriophage T7 specify the beginning and end of the RNA molecule that is transcribed from the DNA. This RNA includes the Spinach sequence, which binds to the chemical DFHBI, causing green fluorescence. The ribosome binding site (RBS) directs the translation of the RNA to produce a fusion protein combining the gene of interest with a short tetracysteine (TC) peptide tag. Additional peptide tags (optional, not shown) can also be included for protein purification. The TC tag binds to a chemical (ReAsH), causing red fluorescence. When performing PERSIA in a microfluidic device, this output is measured by quantitative fluorescence microscopy (b). The amounts of green (c) and red (d) fluorescence measured over time give the user a real-time readout of how much RNA and protein are being produced. Performing the reactions without the DNA-based instructions assesses the nonspecific background signal.
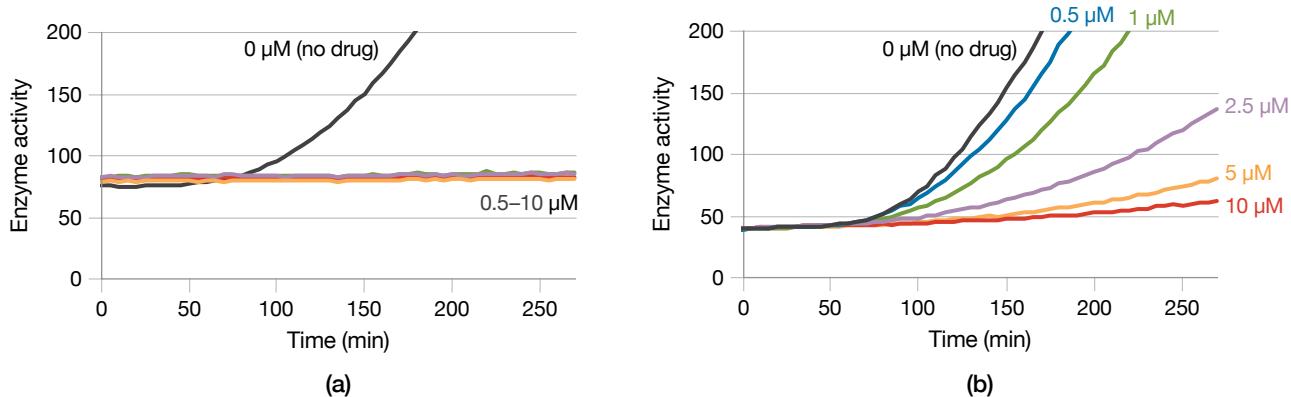
FIGURE 10. The charts illustrate adaptions of PERSIA to assess resistance to antiviral drugs. An additional assay reagent can be employed in a PERSIA cell-free reaction, in this case to measure the activity of the enzyme HIV protease, one of the two major targets for drugs that treat HIV. For the standard reference strain of HIV, the activity of HIV protease is suppressed by low concentrations (0.5 to 10 micromolar, µM) of the drug lopinavir (a). For a genetic variant of the virus known to be resistant to lopinavir, none of the same concentrations tested completely halted protease activity (b).

the protein itself are the signals that turn production on and off. PERSIA employs a short extra sequence (a tetra-cysteine, or TC tag) that adds a few more amino acids to the tail of a protein. The TC tag can then react with the chemical known as ReAsH, resulting in bright red fluorescence. Similarly, the mRNA sequence encoding the protein includes its own tag (named Spinach), which binds to a different chemical, giving a bright green fluorescence. Together these extra components allow us to quantitatively monitor both transcription (RNA production) and translation (protein production) during the reaction itself. In contrast, many other types of analysis would rely on additional experiments to be performed once the cell-free reaction is over. In addition, we have demonstrated that PERSIA can be performed in the very small volumes of a microfluidic device. Our intent is that future versions of the BPU can incorporate PERSIA to produce integrated high-throughput readouts of protein production and function.

We have employed PERSIA to test new design ideas for genetic codes, using the protein readout to tell us whether or not a given genetic code produces a robust quantity of protein. We have also used PERSIA in combination with additional enzymatic assays to probe details of protein structure and function. We have even extended these assays to include measurements of drug activity for different clinically occurring isolates of the HIV protease (Figure 10). Through these approaches, we hope to both accelerate and personalize the effective treatment of viral diseases. Optimistically in the near future, when patients at a clinic or warfighters at a medical field station have their viral infection sequenced, those sequences can be rebuilt as DNA molecules in a BPU-like pipeline and tested with a technique such as PERSIA, and the results can immediately inform clinicians or medics of the best (and personalized) choice of drug regimen for the patients.

Rapid Medical Countermeasures: A Digital There and Back Again

Tools such as PERSIA and the BPU can enable an ambitious vision for fighting pathogens, whether naturally occurring or engineered. Figure 11 shows a pathway that especially leverages the digital nature of DNA sequence information. It relies on onsite sequencing capabilities at the location of concern (such as a remote medical clinic in one part of the world) and BPU-like capabilities at a separate location with more resources.

Consider a newly discovered virus, perhaps in a remote or inaccessible part of the world. Onsite sequencing—such as is now becoming plausible with some next-generation sequencing platforms—could be used to determine the genetic sequence of the virus. That information would then be transmitted to other facilities (centralized or distributed) that would

- Analyze the information content of the sequence to identify which parts of the virus could be targets (such as a viral protease) for a drug or other medical countermeasures (MCMs)

PETER A. CARR, JOHANNA BOBROW, JAMES C. COMOLLI, NICHOLAS J. GUIDO, FRANCES E. NARGI, TODD A. THORSEN, DAVID I. WALSH, MATTHEW E. WALSH, SCOTT T. WICK, AND CATHERINE R. CABRERA
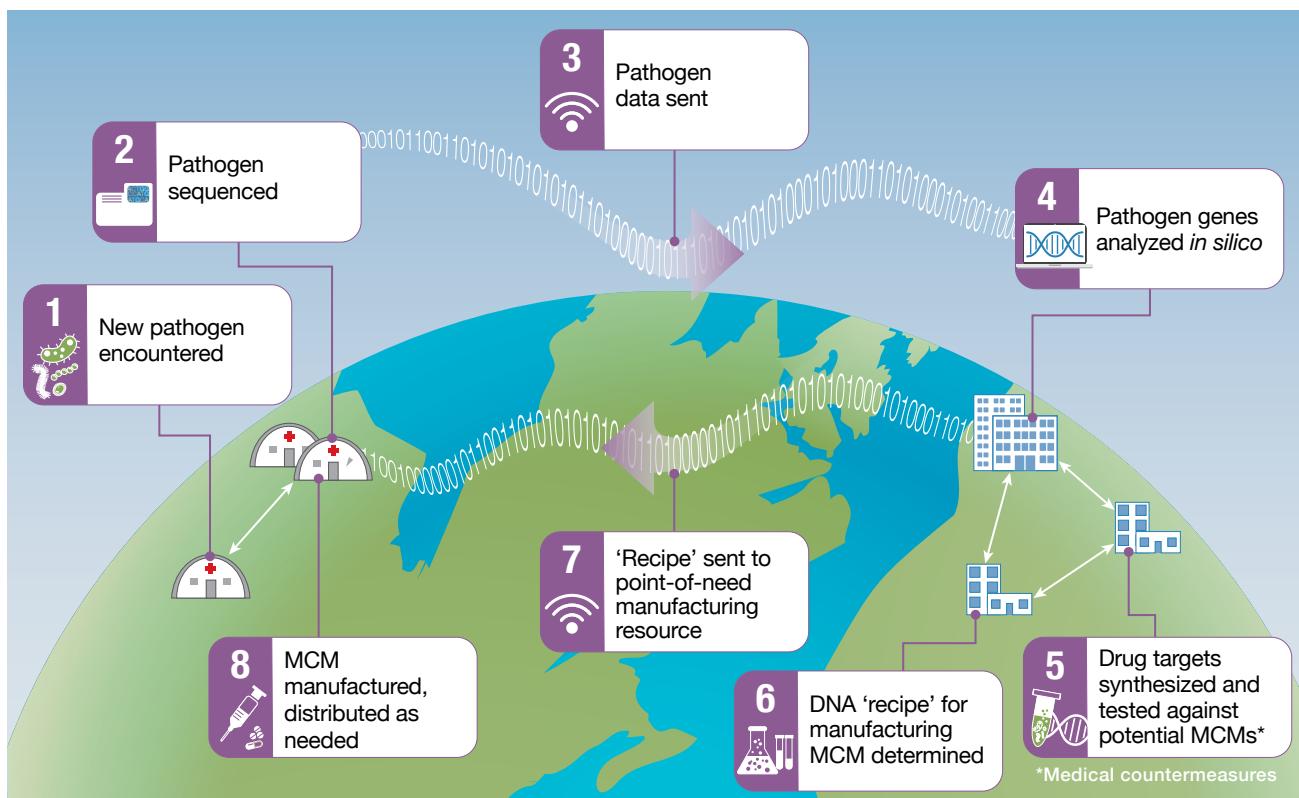
FIGURE 11. The figure illustrates a future pathway for rapid response to new pathogen outbreaks. Sequencing the genetic information of the pathogen allows efforts to commence (anywhere in the world) to determine what medical countermeasures (MCMs) are likely to be effective. A Biomolecular Processing Unit (BPU, see Figure 7) would provide capabilities for steps 5 and 6. In some cases, instructions for making the MCM (such as for a DNA-based vaccine) could also be transmitted back to the point of need for local, distributed manufacturing.

- Reconstruct the physical DNA or RNA molecules encoding those targets
- Employ these DNA sequences to make the corresponding proteins (if the target is a protein and not DNA/RNA itself)
- Perform assays to test whether available known drugs can be repurposed in the hopes of fighting the new threat

The most radical—and perhaps most exciting—possibility for this approach would be if the MCM itself can also be encoded into the information of a DNA sequence and transmitted back to the location(s) where the need is greatest. This pathway would be most plausible if the DNA sequence encodes a protein biologic (e.g., antigen for immunization or therapeutic enzyme), DNA vaccine, therapeutic RNA, or an inert version of the virus itself to be used as a vaccine. Some drugs could also be produced onsite through biosynthesis, with designs for the necessary enzyme genes transmitted from afar.

(Penicillin is one example, requiring only a few enzymes [17].) We have also considered how the same approach could be used to transmit such countermeasure "recipes" through space to remote exploration and colonization missions on other planets [18].

Understanding Technological Risk

Synthetic biology advances are making the engineering of biological systems faster and more straightforward. While this progress is exciting for the wealth of applications enabled, it also raises concerns of possible misuse, both accidental and intentional. For example, the same DNA synthesis technology that enables rapid prototyping could be used to help recreate the (eradicated) smallpox virus from only its digital genetic information. Therefore, one role for the synthetic biology program at Lincoln Laboratory is to assess these technologies and consider long-term consequences for security and safety. Such projects may involve laboratory-based test and

evaluation of these new capabilities or purely analytical studies drawing on the Laboratory's deep expertise and involvement in the synthetic biology research community.

For example, the biotechnology known as CRISPR (for clustered regularly interspaced short palindromic repeats, also CRISPR-Cas or CRISPR-Cas9) has produced a tremendous amount of excitement, not only in research circles but for policy, national security, and society at large. Because of the broad range of applications and an impressive relative ease of use, CRISPR has been quickly adopted around the world and adapted to an impressive array of new applications. CRISPR can be considered a tool (or better yet, a suite of tools) that makes genetic engineering easier. But because CRISPR has received so much excited attention in the press, there is a need to separate real capabilities from hype. However powerful, CRISPR is not a tool that replaces all other tools, nor is it a weapon in and of itself.

Table 2 shows a comparison between CRISPR techniques and previous gene/genome editing approaches. At the heart of each of these is the capacity to programmably home in on an exact location in a DNA molecule (even in a whole genome, which can include billions of possible locations). All of these approaches are programmable—users are able to specify the exact DNA location in the genome they wish to target. Site-directed mutagenesis (SDM) for example requires making the edits in vitro and then transferring the edited material into the cell. In the table, blue regions in the sequence indicate targeting, either with engineered proteins or synthetic short pieces of DNA or RNA. Red (plus green and purple for multiplex automated genome engineering [MAGE]) indicate modifications made to the genome, performed either as a cut and paste operation (zinc finger and TAL effector nucleases [ZFN/TALEN] and CRISPR) or an overwrite operation (SDM and MAGE). CRISPR, on the other hand, achieves its homing function through a short RNA molecule that can be specified and synthesized with relatively accessible and affordable resources.

In each case, the typical next step is to cut the targeted DNA, the start of a genetic cut-and-paste operation. However, the capacity to specifically target a chosen DNA sequence has found many other applications, including engineering new regulatory proteins that turn genes on or off. Furthermore, CRISPR systems have been impressively adaptable for engineering organisms for which previously few tools were available,

Table 2. CRISPR Compared to Earlier DNA Editing Approaches

	ZFN/TALEN	SDM	MAGE	CRISPR-Cas9
NO PROTEIN ENGINEERING REQUIRED	✗	✓	✓	✓
WORKS DIRECTLY WITHIN CELLS	✓	✗	✓	✓
ADAPTS EASILY TO DIFFERENT ORGANISMS	✓	✓	✗	✓

ZFN/TALEN = zinc finger and TAL effector nucleases; SDM = site-directed mutagenesis; MAGE = multiplex automated genome engineering; CRISPR = clustered regularly interspersed palindromic repeats

PETER A. CARR, JOHANNA BOBROW, JAMES C. COMOLLI, NICHOLAS J. GUIDO, FRANCES E. NARGI, TODD A. THORSEN, DAVID I. WALSH, MATTHEW E. WALSH, SCOTT T. WICK, AND CATHERINE R. CABRERA

opening up many new applications and opportunities for scientific discovery. CRISPR has also proven quite facile for engineering the human genome.

Because gene editing technologies such as CRISPR are both powerful and progressing rapidly, concerns have been raised for the potential risks and their impacts on national security. But it is also worth re-emphasizing that CRISPR is a tool. Consider all the things you could suddenly do with a screwdriver if you never had one before. You could potentially build something dangerous or even build a weapon out of screwdrivers, if desired. But rarely would the concern be placed on the screwdriver itself. In a similar vein, with CRISPR, we recommend focusing concerns on specific applications of the tool and not the tool itself. For example, concerns have been raised about where future biothreats will be engineered using the tools of synbio, with some critics speculating that unconventional research environments, such as community biolabs, could be of particular concern. We performed an analysis of several factors impacting different research settings—both traditional (academic, government, industry) and unconventional—and noted that while it is conceivable that a biothreat could be engineered in any such space, community laboratories seem undeserving of special concern (see Table 3).

A recent Lincoln Laboratory effort in synthetic biology biothreat analysis was contributing to the report issued by the U.S. National Academies of Sciences, Engineering and Medicine (NASEM), "Biodefense in the Age of Synthetic Biology" [19]. The committee's work generated a framework for analyzing potential threats enabled by synthetic biology, applied that framework to the current biodefense landscape, and recommended options for risk mitigation. The final report noted three general areas that rose to the highest relative level of concern (among the biothreats considered): (1) recreating known viruses through DNA synthesis; (2) adding capabilities to bacteria by inserting new genetic functions, including genes encoding antibiotic resistance or toxins; and (3) engineering microbes to biosynthesize dangerous chemicals on or in the human body. One of the overarching conclusions of the report was that biodefense efforts focused only on narrow lists of dangerous agents (such as the Select Agent lists) would be insufficient to protect against the range of potential future biothreats enabled by synthetic biology.

Sharing Designs, Parts, and Responsibility

Synthetic biologists have also sought to engineer their own research culture. These efforts are both technical (creating best practices for design, fabrication, and

Table 3. Biothreat Concerns Mapped to Conventional and Unconventional Lab Settings

	LAB SETTING	OVERSIGHT	VISIBILITY	ACCESS TO EQUIPMENT	ACCESS TO TRAINING	CONTAINMENT
TRADITIONAL RESEARCH SETTINGS	Government	Low concern	Moderate concern	High concern	Moderate concern	High concern
	Industry	High concern	High concern	High concern	Moderate concern	Moderate concern
	University	Low concern	Moderate concern	High concern	High concern	Moderate concern
UNCONVENTIONAL RESEARCH SETTINGS	Incubator	Moderate concern	Moderate concern	Moderate concern	Moderate concern	Moderate concern
	Cloud	Moderate concern	Moderate concern	Moderate concern	Low concern	Low concern
	Community	Low concern	Low concern	Low concern	Moderate concern	Low concern
	Personal	High concern	High concern	Low concern	Low concern	Low concern

Legend: ■ Low concern ■ Moderate concern ■ High concern

FIGURE 12. iGEM's annual Giant Jamboree brings together more than 300 teams of synthetic biologists representing upwards of 6,000 participants from dozens of countries. PHOTO CREDIT: IGEM FOUNDATION AND JUSTIN KNIGHT

measurement) and sociological (propagating norms of ethical choices for synbio research, transparency, safety, cooperation, a culture of sharing of designs and parts). Lincoln Laboratory has contributed to a number of these efforts, such as by creating avenues for openness and design sharing of hardware innovations (Metafluidics) and supporting the International Genetically Engineered Machine (iGEM) competition.

iGEM brings together aspiring synthetic biologists from around the world to compete in an annual showcase of their research projects. This "Olympics of Synthetic Biology" not only trains and inspires thousands of students, it also serves as a proving ground for the ideas and ideals of synthetic biology. By providing teams with a large toolkit of standard DNA building blocks (BioBricks) and standards for their construction, iGEM enables these researchers to build their own new genetic designs. Students also produce many of their own new DNA parts, which are then contributed back to the iGEM BioBricks collection, the Registry of Standard Biological Parts [8]. This give-and-get dynamic helps foster a collaborative community of synthetic biology innovators, which grows every year. The yearly cycle of iGEM culminates in the fall Giant Jamboree (Figure 12), where teams convene to share their work, compete for awards, and celebrate each other's achievements.

From simple beginnings at MIT in 2004, the iGEM competition has grown from five U.S.-based teams to 353 teams worldwide in 2019, involving more than 6,000 participants (college, graduate, and high school) each year [20]. The After iGEM organization represents the more than 45,000 iGEM alumni, many of whom have started new synthetic biology companies—often from their own iGEM projects. Lincoln Laboratory staff have been deeply involved in iGEM from the beginning, as part of the first design teams, and have served as team mentors, Jamboree volunteers (including roles of photographer and dance DJ), committee leaders, judges, and the current Director of Judging for the competition. Some foundational DNA parts in the Registry were created by the pre-iGEM 2003 design teams (including a member who is now on the Lincoln Laboratory technical staff) and have been reused by others more than 6,000 times.

Well beyond technical achievement, iGEM also engages its expanding international community with questions of responsible technology development.

PETER A. CARR, JOHANNA BOBROW, JAMES C. COMOLLI, NICHOLAS J. GUIDO, FRANCES E. NARGI, TODD A. THORSEN, DAVID I. WALSH, MATTHEW E. WALSH, SCOTT T. WICK, AND CATHERINE R. CABRERA

Participants are encouraged to consider both how their work affects the world and how the world affects their work. With teams coming from different backgrounds around the world—and representing diverse regulatory and ethical frameworks—iGEM incentivizes strong norms for biosafety, biosecurity, and bioethics. The annual Giant Jamboree draws together specialists from these disciplines and several others (for example, policy, education, public health, biodefense), many of whom serve on its large panel of judges. iGEM becomes a gathering at which these professionals and the student competitors learn from and challenge each other.

In a similar fashion to the sharing of DNA-based designs within iGEM, we at Lincoln Laboratory have sought to foster better adoption and faster advances within the field of microfluidics by enhancing the sharing of hardware designs. We (and others) have noted that new microfluidic technology does not easily make the transition from the lab into common use or commercialization. One major obstacle has been reproducibility. Reproducibility has been hindered by the way microfluidic hardware designs have been incompletely shared in the research literature, with insufficient information transmitted for reproducing someone else's creation. To encourage a more open community of microfluidic innovators, we created Metafluidics, an online resource dedicated to the sharing of microfluidic designs and expertise [15]. Metafluidics provides a way for innovators to share their designs, methods, and operating protocols, so that a different user can try out those innovations. That new user may also remix and redesign, and contribute that modified version back to the Metafluidics community. Figure 13 shows an example of a microfluidic device design hosted at Metafluidics. Since the web resource launched in 2017, this community of users has grown to more than 2,000 members. We continue to support Metafluidics development in collaboration with David Kong (former Lincoln Laboratory technical staff) of the MIT Media Lab through the Living Computing Project funded by the National Science Foundation.

Another hindrance to innovation in microfluidics has been a general requirement for advanced (and expensive) microfabrication facilities, including clean rooms of the type often used for electronic fabrication. We have explored and shared how makerspaces can provide a viable alternative. Makerspaces are dedicated
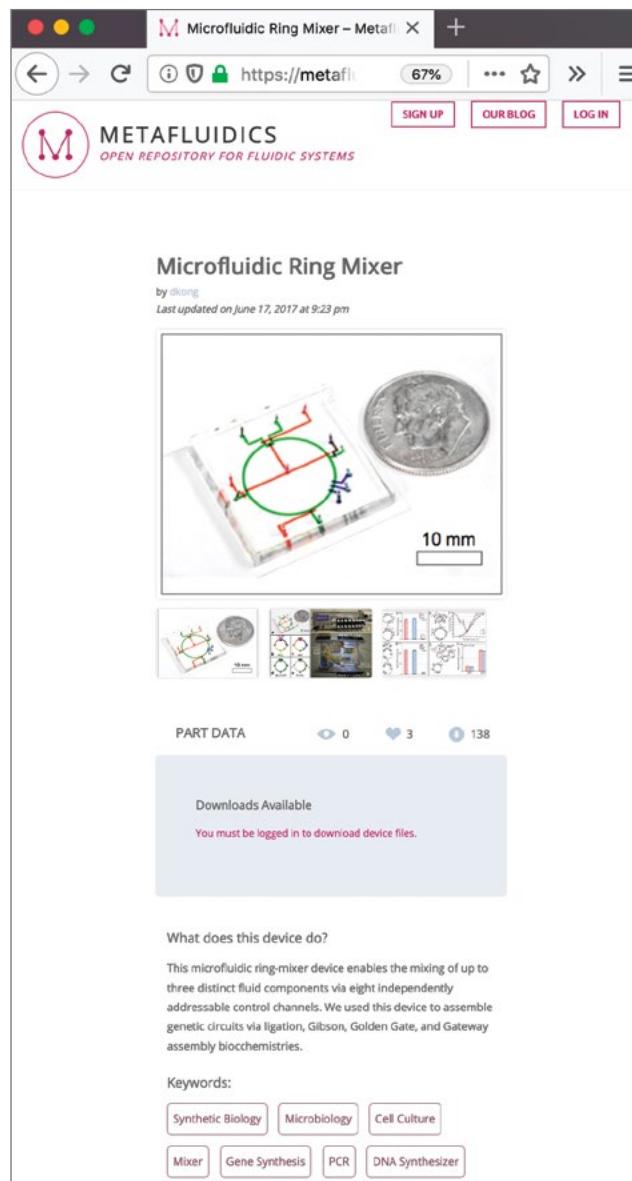
FIGURE 13. This sample page is for a Lincoln Laboratory device hosted on the Metafluidics web resource for shared microfluidic designs.

areas that provide a shared set of tools for a community that wishes to create. The tools available in makerspaces can be diverse, ranging from those found in machine shops and woodworking facilities to those for use in handcrafts and sewing, and more. Many makerspaces have adopted 3D printing technologies and even tools for genetic engineering. (Those that focus mainly on biology and genetic engineering are often called community bio labs, or DIY bio labs.) While the microfluidic devices developed at Lincoln Laboratory have typically required

expensive facilities (e.g., with specialized clean room environments) and equipment, we have also encouraged the use of makerspaces for those wishing to get started [21]. Experiments have included applying 3D printing to create new fluidic devices capable of manipulating and constructing with DNA [15]. We have also encouraged maker communities to contribute and share in solving needs for long-term space exploration and colonization [18]. One exciting application area would be to use a set of (DNA-based) designs to program microbes for biosynthesizing the feedstock materials needed for 3D printing. Thus some types of hardware could be 3D-printed on demand, such as replacement parts for space life-support systems.

Looking Forward

Synthetic biologists look to enable a future in which living things are far easier to engineer and in which such engineering can be applied to help with the world's biggest problems. The potential to engineer crops to withstand the stresses of climate change is one promising area, with the opportunity to alter plants to make them more drought- and heat-tolerant, increase yields, and require less fertilizer [22]. Another key area is improving and protecting human health by addressing rising concerns about naturally occurring pandemics, increasing antibiotic resistance, human accidents, and intentional weaponization of biology. Synthetic biologists are pursuing novel ways to protect against these threats. Engineered biological solutions are increasingly contributing to the bioeconomy [7, 23], and the desire to protect this contribution leads to many questions of how to balance progress with caution.

Lincoln Laboratory combines expertise in synthetic biology and many intersecting disciplines—such as bioinformatics, data science, and electronic device design—to make biology easier to engineer. We seek to enable the growth of the field of synthetic biology, to support national security both broadly and specifically. Broadly, creating new tools for bio-prototyping and measurement can stimulate the activities that fuel the bioeconomy, much in the way access to picks and shovels underpinned the 1849 Gold Rush. More specifically, Lincoln Laboratory's synthetic biology projects address needs in biosecurity by creating hardware and wetware sensor systems to detect specific threats (pathogens, chemical agents) and by accelerating the pursuit of countermeasures against these threats. In the process, we leverage the same diverse expertise to provide advice and analysis to the U.S. government regarding the promise and perils of these new technologies.

Acknowledgments

References

1. Institute of Medicine (U.S.) Committee on Technological Innovation in Medicine, *Sources of Medical Technology: Universities and Industry*, N. Rosenberg, A.C. Gelijns, and H. Dawkins, eds. Washington, D.C.: National Academies Press, 1995.
2. M.B. Elowitz and S. Leibler, "A Synthetic Oscillatory Network of Transcriptional Regulators," *Nature*, vol. 403, no. 6767, 2000, pp. 335–338.
3. T.S. Gardner, C.R. Cantor, and J.J. Collins, "Construction of a Genetic Toggle Switch in *Escherichia coli*," *Nature*, vol. 403, no. 6767, 2000, pp. 339–342.
4. C.J. Paddon, P.J. Westfall, D.J. Pitera, K. Benjamin, K. Fisher, D. McPhee, et al., "High-Level Semi-synthetic Production of the Potent Antimalarial Artemisinin," *Nature*, vol. 496, no. 7446, 2013, pp. 528–532.
5. S. Galanie, K. Thodey, I.J. Trenchard, M. Filsinger Interrante, and C.D. Smolke, "Complete Biosynthesis of Opioids in Yeast," *Science*, vol. 349, no. 6252, 2015, pp. 1095–1100.
6. V.M. Isabella, B.N. Ha, M.J. Castillo, D.J. Lubkowicz, S.E. Rowe, Y.A. Millet, et al., "Development of a Synthetic Live Bacterial Therapeutic for the Human Metabolic Disease Phenylketonuria," *Nature Biotechnology*, vol. 36, no. 9, 2018, pp. 857–864.
7. Engineering Biology Research Consortium, "Engineering Biology: A Research Roadmap for the Next-Generation Bioeconomy," 2019, Retrieved from https://roadmap.ebrc.org.
8. iGEM website, http://igem.org/Main_Page.
9. A.A.K. Nielsen, B.S. Der, J. Shin, P. Vaidyanathan, V. Paralanov, E.A. Strychalski, et al., "Genetic Circuit Design Automation," *Science*, vol. 352, no. 6281, 2016.
10. T. Thorsen, S.J. Maerkl, and S.R. Quake, "Microfluidic Large-Scale Integration," *Science*, vol. 298, no. 5593, 2002, pp. 580–584.
11. T.H. Rider, M.S. Petrovick, F.E. Nargi, J.D. Harper, E.D. Schwoebel, R.H. Mathews, et al., "A B Cell–Based Sensor for Rapid Identification of Pathogens," *Science*, vol. 301, no. 5630, 2003, pp. 213–215.
12. M.S. Petrovick, J.D. Harper, F.E. Nargi, E.D. Schwoebel, M.C. Hennessy, T.H. Rider, and M.A. Hollis, "Rapid Sensors

PETER A. CARR, JOHANNA BOBROW, JAMES C. COMOLLI, NICHOLAS J. GUIDO, FRANCES E. NARGI,
TODD A. THORSEN, DAVID I. WALSH, MATTHEW E. WALSH, SCOTT T. WICK, AND CATHERINE R. CABRERA

for Biological-Agent Identification," *Lincoln Laboratory Journal*, vol. 17 no. 1, 2007, pp. 63–84.

13. Centers for Disease Control and Prevention, Federal Select Agent Program, Select Agents and Toxins, https://www.selectagents.gov/SelectAgentsandToxinsList.html

14. K.S. Boles, K. Kannan, J. Gill, M. Felderman, H. Gouvis, B. Hubby, et al., "Digital-to-Biological Converter for On-Demand Production of Biologics," *Nature Biotechnology*, vol. 35, no. 7, 2017, pp. 672–675.

15. D.S. Kong, T.A. Thorsen, J. Babb, S.T. Wick, J.J. Gam, R. Weiss, and P.A. Carr, "Open-Source, Community-Driven Microfluidics with Metafluidics," *Nature Biotechnology*, vol. 35, no. 6, 2017, pp. 523–529.

16. S. Wick, D.I. Walsh, J. Bobrow, K. Hamad-Schifferli, D.S. Kong, T. Thorsen, et al., "PERSIA for Direct Fluorescence Measurements of Transcription, Translation, and Enzyme Activity in Cell-Free Systems," *ACS Synthetic Biology*, vol. 8, no. 5, 2019, pp. 1010–1025.

17. A.R. Awan, B.A. Blount, D.J. Bell, W.M. Shaw, J.C.H. Ho, R.M. McKiernan, and T. Ellis, "Biosynthesis of the Antibiotic Nonribosomal Peptide Penicillin in Baker's Yeast," *Nature Communications*, vol. 8, no. 15202, 2017.

18. J.E. Snyder, D.I. Walsh, P.A. Carr, and L.J. Rothschild, "A Makerspace for Life Support Systems in Space," *Trends in Biotechnology*, vol. 37, no. 11, 2019, pp. 1164–1174.

19. The National Academies of Sciences, Engineering, and Medicine, *Biodefense in the Age of Synthetic Biology*. Washington, D.C.: The National Academies Press, 2018.

20. "iGEM's Vision: Imaging the Future," iGEM blog, https://blog.igem.org/blog/2019/10/30/igems-vision-imagining-the-future.

21. D.I. Walsh, D.S. Kong, S.K. Murthy, and P.A. Carr, "Enabling Microfluidics: From Clean Rooms to Makerspaces," *Trends in Biotechnology*, vol. 35, no. 5, 2017, pp. 383–392.

22. G.E. Oldroyd and R. Dixon, "Biotechnological Solutions to the Nitrogen Problem," *Current Opinion in Biotechnology*, vol. 26, 2014, pp. 19–24.

23. R.H. Carlson, *Biology is Technology: The Promise, Peril, and New Business of Engineering Life*. Cambridge, Mass.: Harvard University Press, 2010.

About the Authors

Peter A. Carr is a senior staff scientist in the Biological and Chemical Technologies Group at Lincoln Laboratory, where he leads a synthetic biology research program. The central focus of his work is to expand the reach of what can be achieved through engineering living systems—at the level of single genes, entire genomes, or complex communities. He is a founding member of the MIT Synthetic Biology Center, serves as Director of Judging for the International Genetically Engineered Machine (iGEM) competition, and is a coauthor of the National Academies of Sciences, Engineering and Medicine's report *Biodefense in the Age of Synthetic Biology*. Prior to joining Lincoln Laboratory, he led research in DNA synthesis and genome engineering at the MIT Media Lab. He holds a bachelor's degree in biochemistry from Harvard College and a doctorate in biochemistry and molecular biophysics from Columbia University, and studied structural biology for drug discovery as a postdoctoral fellow at the Whitehead Institute for Biomedical Research.

Johanna Bobrow is a member of the technical staff in the Biological and Chemical Technologies Group at Lincoln Laboratory. Over her 20 years at the Laboratory, her research has spanned from sample preparation and technologies for PCR-based detection and identification of biowarfare agents, to protocol development for DNA forensics and sequencing, to physiological status monitoring and hearing damage modeling for warfighter welfare. More recently, she has turned to device design, rapid model prototyping and manufacture of novel devices for biological assays, and synthetic biology research. Prior to joining the Laboratory in 1999, she earned a bachelor's degree in brain and cognitive science at MIT.

James C. Comolli, a member of the technical staff in the Biological and Chemical Technologies Group, joined Lincoln Laboratory in 2015. His research interests include infection therapeutics and diagnostic technologies, biosensors, bacterial communities, pathogen-host interactions, and synthetic biology. Prior to joining the Laboratory, he was the leader of the Biomedical Engineering Group at Draper Laboratory; research director at ECI Biotech, a startup developing microbial diagnostics; and a staff researcher at Johnson & Johnson, working on medical devices. He has a bachelor's degree in biology from Johns Hopkins University and a doctoral degree in cellular and molecular biology from Harvard University. He completed postdoctoral training in bacterial pathogenesis at the University of California, San Francisco, and in microbial physiology at the University of Wisconsin–Madison.

Nicholas J. Guido is a member of the technical staff in the Biological and Chemical Technologies Group. The focus of his work is synthetic biology, including DNA synthesis, genetic networks and pathways, and the application of engineering principles to biological systems in general. Prior to joining Lincoln Laboratory in 2017, he was a senior scientist with Gen9, a start-up DNA synthesis company based in Cambridge, Massachusetts. In this role, he optimized the DNA synthesis process for production of difficult material and created a process for the synthesis of highly variant DNA libraries. In postdoctoral work in the Genetics Department at Harvard Medical School, he studied the modeling of the metabolism of bacteria, leveraging this information to carry out genome engineering and thus increasing the production of high-value products. As a graduate student, he conducted research into the modeling and physical construction of synthetic genetic networks in bacteria. He holds a bachelor's degree in bioengineering from the University of Pennsylvania and a doctorate in bioinformatics from Boston University.

Frances E. Nargi is a technical staff member in the Biological and Chemical Technologies Group. Her research is focused predominantly in the areas of biosensor and biological simulant development. She has worked as a research scientist in the areas of immunology and microbiology since 1985. Her areas of expertise include micro, cell, and molecular biology. She is one of the key team members who developed the CANARY biosensor and has been involved in the evolution and deployment of this sensor for the detection and identification of weapons of mass destruction and plant pathogens. She leads several programs, developed two simulants for use as surrogate biological agents, and designed and built numerous devices to facilitate sample preparation and to support assay development. Prior to joining Lincoln Laboratory in 1999, she worked as a microbiologist for the U.S. Department of Agriculture's Agricultural Research Services, where she established the immunology laboratory in the foot-and-mouth disease unit and evaluated the suitability of potential vaccine candidates. She holds bachelor's and doctoral degrees from the University of Connecticut at Storrs.

Todd A. Thorsen is a technical staff member in the Biological and Chemical Technologies Group, where he pursues research in microfluidics-based engineering, including the design and fabrication of state-of-the-art programmable microfluidic devices, exploitation of the microscale physical properties of fluids (thermal, chemical, optical, and electrical), and the development of microfluidic platforms for medicine. Prior to joining Lincoln Laboratory, he was a faculty member in the MIT Department of Mechanical Engineering. He has authored or coauthored 34 papers and two book chapters, has 40 issued and pending patents in the field of microfluidics, and has more than 8,800 citations of his published works. He holds a bachelor's degree in biology from the University of California, San Diego; a master's degree in infectious disease from the University of California, Berkeley; and a doctoral degree in biochemistry and molecular biophysics from the California Institute of Technology.

David I. Walsh III is a member of the technical staff in the Biological and Chemical Technologies Group. His research is focused on developing in vitro models and rapid, autonomous biosensors by using advanced 3D-printing techniques. Prior to joining Lincoln Laboratory in 2015, he was a National Science Foundation Graduate Research Fellow at Northeastern University, a visiting scholar at KTH Royal Institute of Technology in Stockholm, and a co-op student at Sandia National Laboratories in the Biosystems Research and Development group. He holds a bachelor's degree in nuclear engineering from Rensselaer Polytechnic Institute and a doctorate in bioengineering from Northeastern University. His doctoral dissertation on point-of-care diagnostics focused on the use of makerspaces to democratize the development of point-of-care biomicrofluidic devices beyond advanced engineering laboratories.

Matthew E. Walsh is an associate technical staff member in the Biological and Chemical Technologies Group. Since joining Lincoln Laboratory in 2014, he has worked on a range of programs supporting synthetic biology efforts. His technical work often relies on applying (bio)analytical methods to Department of Defense–specific applications. He is the principal investigator for the Rapid Medical Countermeasures program that looks to apply machine learning to therapeutic antibody development. He is a member of the 2018 Class of the Emerging Leaders in Biosecurity Initiative and holds a bachelor's degree in chemistry from Skidmore College.

Scott T. Wick is an associate staff scientist in the Biological and Chemical Technologies Group at Lincoln Laboratory. His current research efforts focus on design and integration of molecular, cellular, and synthetic biology into medical and environmental sensor platforms built to detect pathogenic organisms, toxic chemicals, and important biomarkers related to health and human performance. Prior to joining the Laboratory in 2004, he

PETER A. CARR, JOHANNA BOBROW, JAMES C. COMOLLI, NICHOLAS J. GUIDO, FRANCES E. NARGI, TODD A. THORSEN, DAVID I. WALSH, MATTHEW E. WALSH, SCOTT T. WICK, AND CATHERINE R. CABRERA

was a research staff member at GPC Biotech (formerly Mitotix, Inc.), developing protein-based anticancer treatments, and at Sterling-Winthrop pharmaceuticals, developing antibody-based anticancer therapies. He holds a bachelor's degree in biotechnology from the Rochester Institute of Technology and a master's degree from Harvard University, publishing a thesis developing novel protein chimeras for vascular disease therapy.

Catherine R. Cabrera is the leader of the Biological and Chemical Technologies Group at Lincoln Laboratory. She joined the Laboratory in 2002, initially working on hardware and software development for the identification of biowarfare agents. She was part of the team that received an R&D 100 Award for the development of the PANTHER automated cell-based bioaerosol sensor, which has since transitioned to operational use for building protection and plant pathogen detection. She currently oversees a diverse portfolio of programs that include ones on molecular biomarkers for heath and performance, advanced DNA forensics, and engineered and synthetic biology. Her areas of technical expertise include microfluidics; biodefense technologies, systems, and architectures; red/blue team analysis; microbiome and human health; point-of-need diagnostics; genetic and epigenetic biomarkers of health and activity; and use of physiological status indicators to provide early warning of exposure to chemical warfare agents or pathogens. She holds bachelor's degrees in biochemistry and chemical engineering from Rice University and a doctorate in bioengineering from the University of Washington. Her doctoral research was focused on developing fieldable technologies to detect pathogens in resource-limited environments.

In Lincoln Laboratory's Sensorimotor Technology Realization in Immersive Virtual Environments (STRIVE) Center is a 24-foot virtual reality dome, the Computer Assisted Rehabilitation Environment (CAREN), which allows users to experience immersion in a simulated world. A user interacts with an environment that is displayed on a 360-degree screen while walking on a 6-degrees-of-freedom motion platform that mimics the environment's terrain. A motion-capture system measures and analyzes the user's movements in real time. In the image above, the CAREN is being used to monitor how an individual interacts with a prototype exoskeleton that is designed to help a person hike with less strain on lower-limb muscles. The STRIVE Center is also used to study patients' cognitive and physical performance, test rehabilitation techniques, and assess the effects of training regimens. More about the STRIVE Center can be found in the appendix to the article "Biomechanical Sensing and Algorithms" on page 165.