TIME-SCALE MODIFICATION OF COMPLEX ACOUSTIC SIGNALS

T.F. Quatieri1  R.B. Dunn1  T.E. Hanna2

1MIT Lincoln Laboratory, Lexington, MA, USA
2NSMRL, Box 900, Naval Submarine Base New London, Groton, CT, USA

ABSTRACT
A new approach is introduced for time-scale modification of short-duration complex acoustic signals to improve their audibility. The technique constrains the modified signal to take on a specified spectral characteristic while imposing a time-scaled version of the original temporal envelope. Both full-band and sub-band representations of the temporal envelope are considered. In the full-band case, the modified signal is obtained by appropriate selection of its Fourier transform phase. In the sub-band case, using locations of maxima in the sub-band temporal envelopes, the phase of each bandpass signal is formed to preserve “events” in the envelope of the composite signal. The approach is applied to synthetic and actual short-duration acoustic signals consisting of closely-spaced and overlapping sequential time components.

1. INTRODUCTION
Short-duration complex sounds, as from the closing of a stapler or the tapping of a drum stick, often consist of a series of brief components which are closely spaced as well as overlapping in time, and are therefore difficult to discern. This class of sounds is characterized by a rich temporal envelope structure which appears to play an important role in auditory discrimination. In many nonspeech sounds [3] [4] [5], as well as unvoiced speech sounds [10], the temporal envelope has been found to be a perceptual cue for recognition. Enhancement of the audibility of these sounds should then capitalize on the perceptual importance of the temporal envelope.

In one approach to signal enhancement, sometimes referred to as “slow-motion audio replay”, a signal is time-scaled expanded without altering its spectral character, thus allowing the listener to capture the sequential components of short-duration complex signals. Methods of time-scale modification, however, even high quality techniques developed in the speech context such as the phase vocoder [2] and sine-wave analysis/synthesis [8], can smear the fine structure of the temporal envelope. Loss of temporal resolution results from windowing in analysis which can dull the sound and merge components, and from phase dispersion in synthesis which can introduce a reverberant, tinny quality. In this paper, an approach is proposed for time-scale modifying a complex acoustic signal which preserves the fine structure of the signal's temporal envelope.

The outline of the paper is as follows. Section 2 presents an iterative algorithm for reconstruction of a signal from its temporal and spectral envelopes. In section 3, the results of section 2 are used to time-scale modify a signal based on a modified temporal envelope. In section 4, a sub-band signal representation forms the basis of an alternate method which exploits the maxima of the temporal envelope in different bands. Section 5 summarizes and describes current efforts.

2. RECONSTRUCTION FROM TEMPORAL AND SPECTRAL ENVELOPES
Under certain conditions a signal can be represented by a partial specification either in the time domain, the frequency domain, or both domains. For the purpose of time-scale modification, the temporal and spectral envelopes are specified. The spectral envelope is defined as the magnitude of the Fourier transform. The temporal envelope is defined to reflect the distinctive “events” of the complex acoustic signal; e.g., the start and stop time of a click or the modulation pattern of two beating sine waves. One definition of temporal envelope, used typically in the context of bandpass signals, relies on the Hilbert transform; a different definition uses measurements of attack and decay dynamics. The advantage of the former approach is that it allows the development of conditions for representation of a signal by its temporal and spectral envelope, while the advantage of the latter approach is that the temporal envelope can be better controlled to reflect specific events [9].

Consider, in particular, the definition of temporal envelope based on the Hilbert transform. A discrete-time signal \( x(n) \) is given in analytic form by

\[
s(n) = x(n) + j\hat{x}(n) \quad (1a)
\]

where \( \hat{x}(n) \) is the Hilbert transform of \( x(n) \) and which is written in polar form as

\[
s(n) = a(n)\exp[j\phi(n)] \quad (1b)
\]

where \( a(n) \) is the temporal envelope of the sequence. The Fourier transform of (1a) is given by

\[
S(\omega) = X(\omega) \quad 0 \leq \omega \leq \pi \quad (2a)
\]

\[
S(\omega) = X(\omega) \exp[j\phi(\omega)] \quad (2b)
\]

and which is written in polar form as

where \( A(\omega) \) is the spectral envelope of the sequence.

I-213

0-7803-0946-4/93 $3.00 © 1993 IEEE
For a large class of sequences \( x(n) \), reconstruction can be performed iteratively from the temporal and spectral envelopes, \( a(n) \) and \( A(\omega) \), respectively [9]. In one iterative algorithm, the desired temporal envelope is imposed in the time domain via the Hilbert transform, and the specified spectral envelope is imposed in the frequency domain via the Fourier transform. Specifically, on the \( k \)th iteration of the algorithm, the analytic form of the signal estimate \( \hat{x}_k(n) \) is obtained in polar form as \( \hat{s}_k(n) = a_k(n)e^{j\hat{\phi}_k(n)} \). The temporal envelope of this estimate is replaced by the desired function \( a(n) \)

\[
\hat{s}_k(n) = a(n)e^{j\hat{\phi}_k(n)}
\]

The spectral envelope of the Fourier transform of \( \hat{s}_k(n) \), \( \hat{A}_k(\omega)e^{j\hat{\theta}_k(\omega)} \), is then replaced by the specified spectral envelope

\[
\hat{S}_k(\omega) = A(\omega)e^{j\hat{\theta}_k(\omega)}
\]

The successive iterate \( \hat{x}_{k+1}(n) \) is formed by taking the real part of the inverse Fourier transform of (4). In practice, the continuous Fourier transform is replaced by a discrete Fourier transform (FFT) which is sufficiently long to avoid aliasing.

Empirically, it was found that the mean-squared error in the temporal envelope, \( \sum_n [a_k(n) - a(n)]^2 \), is nonincreasing as a function of iteration. For a variety of signals tested, and with an initial condition set to the desired temporal envelope, convergence was achieved within a few hundred iterations. An example of reconstructing a sequence from its temporal and spectral envelopes is illustrated in Figure 1. The synthetic signal consists of the sum of two damped sine waves of different frequencies displaced by 100ms which was selected to resemble a response from a closing stapler. The mean-squared error in the temporal envelope as a function of iteration is also shown in Figure 1.

3. TIME-SCALE MODIFICATION USING THE FULL-BAND ENVELOPE

The essence of the approach to time-scale modification, given a specified spectral envelope, is to select a Fourier-transform phase that results in a sequence with a time-scaled version of the original temporal envelope. A close match of both a specified spectral envelope and modified temporal envelope may, however, not be "consistent" with the relationship between a sequence and its Fourier transform. Consider, for example, an exponentially damped sine wave. Expansion of the temporal envelope by slowing its rate of decay will narrow the signal's resonant bandwidth; arbitrarily slowing the decay rate and maintaining the original resonant bandwidth may violate constraints on the signal's time and frequency concentrations [7]. Consequently, in general there may not necessarily exist a sequence jointly satisfying temporal and spectral envelope constraints. The signal modification problem then is formulated as finding a Fourier-transform phase that results in a sequence whose

\[ E = \sum_n [\hat{s}(n) - a(n)]^2 \]

subject to the constraint that the spectral envelope of the modified sequence equals \( A(\omega) \) or equals a version of \( A(\omega) \) which is more "consistent" with \( a(n) \) but which does not compromise the "spectral character" of the original sequence [9]. The iterative reconstruction algorithm of the previous section represents a suboptimal approach to solving this nonlinear optimization problem.

An example is illustrated in Figure 2 where the sequence in Figure 1 has been time-scale expanded by a factor of two using 300 iterations. The desired temporal envelope \( a(n) \) was obtained by upsampling the original temporal envelope \( a(n) \). The spectral envelope was left intact except for a mild
spectral (resonant) sharpening of $A(\omega)$ to improve "consistency" with $n(\cdot)$. The mean-squared error between the desired and estimated temporal envelopes in this case does not approach zero. Nevertheless, the two closely-spaced and overlapping components of the signals, which were barely audible in the original signal, were perceptually distinct in the modification.

4. TIME-SCALE MODIFICATION USING SUB-BAND ENVELOPES

A second approach to time-scale modification relies on a filterbank signal representation. The output of each filter is viewed as an amplitude and phase modulated sine wave whose amplitude and unwrapped phase is interpolated to perform time-scale modification. This technique, which is the approach to time-scale modification used in the phase vocoder, generally fails for short-duration complex signals of the form in Figure 1. The limitation stems from the filter lengths which are on the order of 10-20 ms for adequate spectral resolution, and from the lack of control on the phase relation among filters [2]. Although with interpolation, the shape of the temporal envelope of each filter can be preserved, the envelope of the composite signal may differ significantly from the original.

An important element of the new approach is to design the filter bank so that each filter output reflects distinctive “events” which characterize the temporal envelope of the input signal. This constraint requires filters of short duration. The filters must also be smooth in time so that the temporal envelope of each filter output does not exhibit spurious peaks due to the filter itself. To meet these requirements a perfect reconstruction filterbank, with twenty one uniformly spaced filters, was designed using a prototype Gabor filter with an effective duration of about one millisecond. The second (related) element of the new approach is that the temporal envelope of the time-scaled composite output is controlled by manipulation of the phase of each channel. The phase control uses the locations of maxima of the channel temporal envelopes which are assumed to occur at the time of the events of the input signal. This approach was inspired by the work of Mallat in a wavelet representation of waveform “singularities” [6] and also by an approach to time-scale modification of voiced speech which is based on glottal onset times [8].

To develop the time-scale modification algorithm more specifically each bandpass Gabor filter, denoted by $h_k(n)$, is assumed to be complex so that the temporal envelope of the output of the $k$th channel can be written as

$$e_k(n) = |y_k(n)|$$

where $y_k(n)$ is the convolution of the bandpass filter with the input sequence

$$y_k(n) = x(n) * h_k(n)$$

The phase of each bandpass output is given by

$$\theta_k(n) = \tan^{-1}(Im[y_k(n)]/Re[y_k(n)])$$

and is assumed to be unwrapped. The occurrence time of an event within each channel is defined as the location of the maximum of $e_k(n)$ and will be denoted by $n_k(k)$. (Only one occurrence time is assigned for each channel, but more generally multiple occurrence times will be required.) An interesting property of the occurrence times $n_k(k)$ is that they tend to cluster near singularities such as sharp attack or decay times. This observation is similar to that made by Mallat for nonuniform filterbanks [6]. An example of this clustering property will be illustrated below.

With time-scale modification by a factor of $q$, the modified filter output is given by

$$\tilde{y}_k(n) = \hat{e}_k(n) \cos[q \hat{\theta}_k(n)]$$

where $\hat{e}_k(n)$ and $\hat{\theta}_k(n)$ denote the channel envelope and unwrapped phase, respectively, interpolated by a factor of $q$. Since the original phase relation among channels is lost through the phase modification in (8), the temporal envelope of the composite signal $\sum_k \tilde{y}_k(n)$ will generally not be a time-scaled version of the original signal envelope. A phase correction is therefore introduced in each channel which makes the phase of the modified filter output $\tilde{y}_k(n)$ at time $q n_k(k)$ equal to the phase at the event occurrence time in the original time scale. Since the values of $n_k(k)$ tend to cluster near singularities, the original phase relation among channels will be approximately preserved at these event occurrence times. Denoting the phase correction by $\phi_k$, the modified channel signal becomes

$$\tilde{y}_k(n) = \hat{e}_k(n) \cos[q \hat{\theta}_k(n) + \phi_k]$$

where

$$\phi_k = \theta_k(n_k(k)) - q \theta_k(n_k(k))$$

Given the clustering property of $n_k(k)$, a refinement of the procedure modifies $n_k(k)$ to take on a single value in each cluster, whereby a group of filter outputs will contribute to a specific event by way of its phase relations.

I-215
Figure 4: Time-scale expansion of a response from a closing stapler; (a) Original; (b) Expansion.

To illustrate the filterbank approach, Figure 3 shows an example where the signal in Figure 1 has been passed through the Gabor filterbank with the phase correction of (9b). Figure 3a shows the maxima of the temporal envelope of the filter outputs as a two-dimensional function of time and filter number, and demonstrates that the maxima cluster near the onset times of the two damped sine waves. Consequently, the occurrence times $n(k)$ were altered to take on the average of one or the other of the two clusters. The assignment was based on the smallest distance to each cluster average. The signal time-scale expanded by a factor of two, shown in Figure 3b, approximately preserves a time-scaled version of the original temporal envelope with excellent time resolution. For reference, the time-scale expanded signal without phase correction, illustrated in Figure 3c, severely distorts the original temporal envelope. The sub-band approach was also applied to the signals referred to in the introduction of this paper. Figure 4 shows the modification of an actual acoustic signal from a closing stapler, while Figure 5 illustrates the modification of a sequence from a percussion instrument.

In these examples, as well as with a variety of other signals, the components of the acoustic signal, which were barely audible in the original, are enhanced through time expansion of the temporal envelope. Also observed was a narrowing of resonant bandwidths which manifest itself perceptually as a "sharpening" of the sound. Given the fidelity of the expanded temporal envelope and given constraints on a signal's time and frequency concentrations [7], this effect is not unexpected.

5. DISCUSSION

Full-band and sub-band approaches to time-scale modification of complex acoustic signals were described. The methods preserve the time-scaled temporal envelope of a signal and for enhancement capitalize on the perceptual importance of a signal's temporal structure. In addition to a more rigorous development of the two methods, as well their relative advantages, a number of other areas are being considered. One important problem is the processing of long waveforms which will require short-time analysis. In the full-band approach, short-time Fourier transform analysis should be followed by a smooth splicing of the time-scaled segments. In using the sub-band approach over a long duration, a more challenging problem is the short-time representation of multiple events within each channel, in contrast to a single event as assumed within this paper.

REFERENCES