
Abstract

An automatic technique for estimating and modeling the glottal flow derivative source waveform from speech, and applying the model parameters to speaker identification, is presented. The estimate of the glottal flow derivative is decomposed into coarse structure, representing the general flow shape, and fine structure, comprising aspiration and other perturbations in the flow, from which model parameters are obtained. The glottal flow derivative is estimated using an inverse filter determined within a time interval of vocal-fold closure that is identified through differences in formant frequency modulation during the open and closed phases of the glottal cycle. This formant motion is predicted by Ananthapadmanabha and Fant to be a result of time-varying and nonlinear source/vocal tract coupling within a glottal cycle. The glottal flow derivative estimate is modeled using the Liljencrants-Fant model to capture its coarse structure, while the fine structure of the flow derivative is represented through energy and perturbation measures. The model parameters are used in a Gaussian mixture model speaker identification (SID) system. Both coarse- and fine-structure glottal features are shown to contain significant speaker-dependent information. For a large TIMIT data subset, averaging over male and female SID scores, the coarse-structure parameters achieve about 60% accuracy, the fine-structure parameters give about 40% accuracy, and their combination yields about 70% correct identification. Finally, in preliminary experiments on the counterpart telephone-degraded NTIMIT database, about a 5% error reduction in SID scores is obtained when source features are combined with traditional mel-cepstral measures.