Hirak Dasgupta, “Detection of excitation epochs and voicing in speech signals using Hilbert envelope,” *Ph.D. Thesis,* Department of Electrical Engineering, Indian Institute of Technology Bombay, September 2022 (Supervisor: Prof. P. C. Pandey).

Abstract

The voiced speech signal is characterized by quasi-periodic impulsive excitations, which occur around the glottal closure instants and are known as the excitation epochs. The research objective is to develop an excitation epoch and voicing detection technique suitable for normal speech, telephone-quality speech, and speech with voice disorders. For this purpose, investigations are carried out to develop a technique using Hilbert envelope of the speech signal and employing single-pass processing with low algorithmic delay and computational requirements. The evaluations are carried out using parallel speech and EGG signals and the ground truths obtained from the EGG signal.

An excitation epoch detection technique using Hilbert envelope of the speech signal to enhance the impulsive excitations is presented. It comprises dynamic range compression, squared Hilbert envelope calculation, saliency enhancement of the excitation epochs, and epoch marking. The dynamic range compression reduces the amplitude variation of the input signal, and the saliency enhancement comprising dynamic peak detection, nonlinear smoothing, and differentiation suppresses the residual ripples related to the vocal-tract filter response. Two epoch marking methods are used. The first method uses amplitude-duration thresholding and is computationally simple. The second method uses detection of maximum-sum subarray peaks, provides a self-correcting epoch marking, and is better suited for speech characterized by high jitter and shimmer and varying spectral tilt. A voicing detection technique using the epochs detected by the Hilbert envelope-based epoch detection and an inter-epoch similarity measure is presented. It comprises frame segmentation of the squared Hilbert envelope, calculation of inter-epoch similarity measure as normalized covariance of Hilbert envelope of the first two inter-epoch intervals in a frame, and voicing decisions based on thresholding followed by a median filter for suppression of isolated detections. The excitation epoch detection and voicing detection techniques are integrated with modifications to obtain an excitation epoch and voicing detection technique with low processing delay and computational requirements.

The technique employs single-pass processing with an algorithmic delay of less than 60 ms. It uses a buffer of length equal to the longest pitch period for epoch marking. The technique was evaluated using databases with simultaneously acquired speech and EGG signals: ‘CMU-ARCTIC database’ for normal speech and ‘Saarbruecken voice database’ for speech with voice disorders. The averaged accuracy-weighted identification rates of the excitation epoch detection for normal speech signals, telephone-quality speech signals, and speech signals with voice disorders were 80.49%, 79.85%, and 70.52%, respectively, and the corresponding averaged voicing decision errors were 9.26%, 9.86%, and 14.52%, respectively. The performance measures for excitation epoch and voicing detections compared favorably with the earlier techniques.