A. R. Jayan / Supervisor: Prof. P. C. Pandey, “Enhancement of speech intelligibility using acoustic properties of clear speech,” *Ph.D. Thesis*, Department of Electrical Engineering, Indian Institute of Technology Bombay, May 2014.

Abstract

Speech intelligibility can be enhanced using acoustic properties of “clear speech”, the speech produced by a speaker with an intention to improve intelligibility in a difficult communication environment. The research objective is to devise a signal processing technique based on the properties of clear speech for improving perception of stop consonants for use in speech communication devices and hearing aids. The method assumes clean speech to be available and processing is performed to make it robust towards further degradations under adverse listening conditions.

Modification of speech characteristics around acoustic landmarks, the regions with a concentration of acoustic cues, is expected to improve speech intelligibility. Detection of landmarks associated with stop consonants was investigated using (i) subband energies and centroids, (ii) parameters from Gaussian mixture modeling, (iii) spectral moments, and (iv) spectral moments with tone-addition. Comparing the algorithmic and computational delays involved in landmark detection, rate of change of spectral centroid derived from tone-added speech spectrum was identified as the parameter most suited for real-time detection of burst onset landmarks. Automated enhancement of speech intelligibility by consonant-vowel ratio (CVR) modification and time-scale modification was investigated. CVR modification involved amplification of vowel-to-consonant (VC) and consonant-to-vowel (CV) transition segments by 9 dB. Sinusoidal model based analysis-synthesis was used for time-scale modification of CV transition segment by a factor of 1.5. Listening tests were conducted on normal-hearing subjects using isolated VCV utterances with speech-spectrum shaped noise as a masker. CVR modification improved recognition scores by 7, 18, and 25% at SNRs of 0, −6, and −12 dB, respectively. No statistically significant improvements were obtained for time-scale modification.

Further investigations were performed on CVR modification using a real-time compatible algorithm. Test material involved utterances with CV syllables, VC syllables, and MRT wordlists. Listening tests were conducted using normal-hearing subjects with speech-spectrum shaped noise as the masker. The improvements in recognition scores for the CV syllables were 8, 9, and 19% at SNRs of 0, −6, and −12 dB, respectively. The corresponding improvements were 9, 11, and 14% for VC syllables. For MRT wordlists, the corresponding improvements were and 8, 9, and 11%, and the improvements were equivalent to an SNR advantage of 3 dB. There was no significant increase in the response times for the processed stimuli indicating no significant increase in the perceptual load. The technique for CVR modification was implemented on a DSP board based on a 16-bit fixed point processor with on-chip FFT hardware and tested for satisfactory real-time operation. Thus the investigations have shown that CVR modification using the proposed technique may be used for improving speech perception under adverse listening conditions.