

Cepstral Liftering Techniques for Voice Quality Assessment

Specific regions of the cepstrum of voiced speech are selected for analysis with a view to determining a harmonics-to-noise ratio (HNR) estimate. Firstly, the low quefrequency portion of the cepstrum is analysed. The Fourier transformed liftered cepstrum approximates a noise baseline from which the harmonics-to-noise ratio is estimated. The present study highlights the manner in which the cepstrum-based noise baseline estimate is obtained, essentially behaving like a moving average filter applied to the power spectrum for voiced speech. As such, the noise baseline, which is taken to approximate the noise excited vocal tract, is shown to be influenced by the window length and the shape of the glottal source spectrum. Two approaches (a harmonic pre-emphasis technique and a symmetric baseline technique) are implemented to overcome the glottal source and window length dependences. The results indicate accurate HNR estimation using the new methods.

The high quefrequency region of the cepstrum has been investigated with a view to providing a correlate of voice quality. Specifically, the high quefrequency region is characterised by rahmonics peaks spaced at the pitch period and its sub-multiples. It is known that the amplitude of the first rahmonic, R1 has a correspondence with the richness of the harmonic spectrum for voiced speech, however a formal description has, to date, remained absent. The present study provides a theoretical description of rahmonic analysis of voiced speech containing aspiration noise and hence derives a definition for R1. It is shown that R1 is proportional to a geometric mean harmonics-to-noise ratio (gmHNR), where the gmHNR is defined as the mean of the individual spectral (i.e. at specific frequency locations) harmonics-to-noise ratios in dB. The technique is tested using synthesized voice signals. R1 is found to be sensitive to all forms of waveform aperiodicities and is shown to depend on analysis window length and fundamental frequency, f_0 . A pitch-synchronous, harmonic-limited spectral analysis is implemented to alleviate the window length/ f_0 dependence of the measure. A discussion of the results of previous studies employing R1 in the analysis of human voice signals is given in light of the definition and present results on synthesis.