Evaluation of cepstrally derived harmonics-to-noise ratios in voice signals

Olatunji Akande & Peter Murphy

Department of Electronic and Computer Engineering
University of Limerick, Limerick, Ireland.
{peter.murphy; olatunji.akande}@ul.ie

Abstract

The harmonics-to-noise ratio (HNR) has been shown to provide a useful, non-invasive index for assessing and analyzing voice disorders. As such, a valid and reliable method for estimating levels of noise and hence HNR, in human voice is important for effective evaluation and management of voice pathology. An objective method to determine the accuracy of two cepstrum-based techniques described in the literature for calculating the harmonics to noise-ratio (HNR) in voiced speech is proposed. In order to draw an objective comparison between the two HNR estimation techniques, the same analysis parameters (analysis window size, voice source parameters (same open and closed quotient), same level of noise for each synthetic speech signal being analyzed) are used. A reference benchmark is defined for HNR estimation comparison as the ratio of the clean signal energy to the noise energy expressed in decibel. This reference HNR forms the basis upon which the two HNR techniques are evaluated for accuracy. Accuracy of a method is measured in terms of deviation of the estimated HNR from the reference HNR for a given noise level. A comprehensive explanation of the origins of the differences in accuracy of each approach is presented.

1. Introduction

The cepstrum is used to estimate the harmonics-to-noise ratio (HNR) in speech signals [1], [2]. The basic procedure presented in [1] is as follows; the cepstrum is produced for a windowed segment of voiced speech. The rahmonics are zeroed and the resulting liftered cepstrum is inverse Fourier transformed to provide a noise spectrum. After performing a baseline correction procedure on this spectrum (the original noise estimate is high), the logarithm of the summed energy of the modified noise spectrum is subtracted from the logarithm of the summed energy of the original harmonic spectrum in order to provide the harmonics-to-noise ratio estimate (Fig.1),
A modification to this technique, [2], illustrates problems with the baseline fitting procedure and hence does not adjust the noise baseline but calculates the energy and noise estimates at harmonic locations only (Fig.2). In addition, rather than zeroing the rahmonics the cepstrum is low passed filtered to provide a smoother baseline (the reason the baseline shifting is not required is due to the window length used).

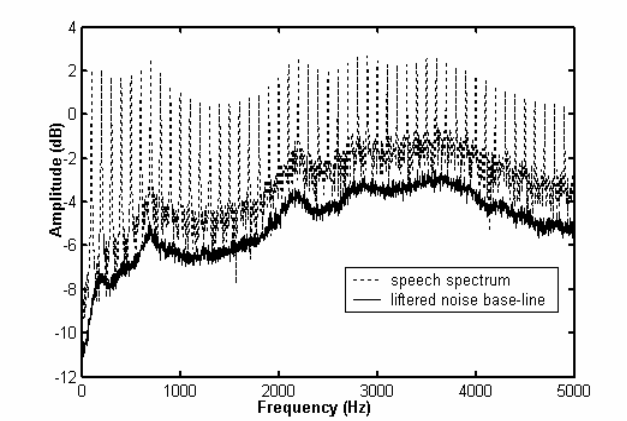

Fig.1. HNR estimation using de Krom [1] cepstral baseline technique using 1024, 2048, 4096 window lengths (1024 shown).

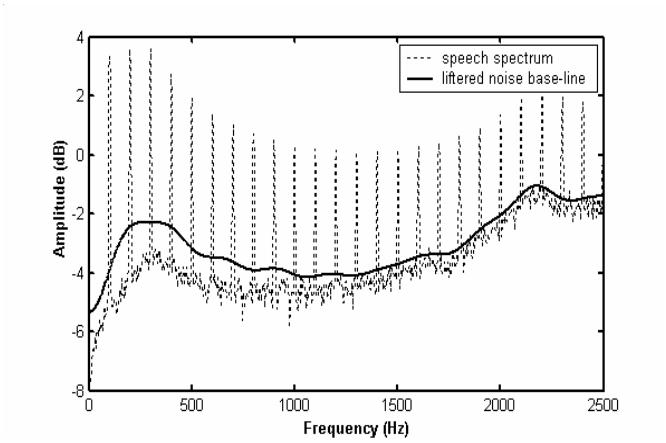

Fig.2. HNR estimation using Qi and Hillman [2] cepstral baseline technique (window length 3200 points).

Each of these approaches, [1], [2], provide useful analysis techniques and data for studies of voice quality assessment, however, to date, neither method has been tested on synthesis data with *a priori* knowledge of the harmonics-to-noise ratio. The present study uses known amounts of random noise added to the glottal source to systematically test, and evaluate the accuracy of the two cepstrum based methods

2. Method

In order to evaluate the performance of the existing cepstrum-based HNR estimation techniques, synthesized glottal source and vowel /AH/ waveforms are generated at five fundamental frequencies (f0s) beginning at 80 Hz increasing in four steps of 60 Hz up to 320 Hz, covering modal register. The model described in [3] is adopted to synthesize the glottal flow waveform while the vocal tract impulse response is modeled with a set of poles. Lip radiation is modeled by a first order difference operator $R(z)=1-z^{-1}$. A sampling rate of 10 kHz is used for synthesis. Noise is introduced by adding pseudo-random noise to the glottal pulse via a random noise generator arranged to give additive noise of a user-specified variance (seven levels from std. dev. 0.125%, doubling in steps up to 2 %). The corresponding HNRs for the glottal flow waveform are 58 dB to 22 dB, decreasing in steps of 6 dB.

2.1 Results

The HNR plotted against f0 is shown for (a) deKrom [1] (Fig.3) (b) Qi and Hillman [2] (Fig.4) for glottal source waveform. In order to evaluate the performance of a method, the estimated HNR is compared to the original HNR (dotted curve) in the figures. It can be seen from Figs.3 & 4 the estimated HNRs deviate from their true values for each method. Method described in [1] overestimates the HNR while [2] underestimates it.

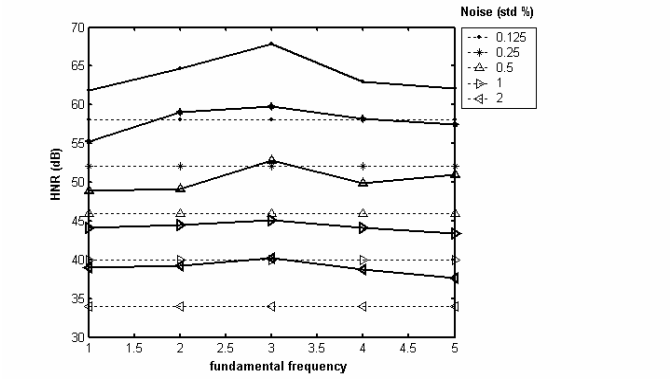

Fig.3. Estimated HNR (solid line, de Krom [1]) versus f0 for synthesized glottal source waveforms (dotted line – actual HNR).

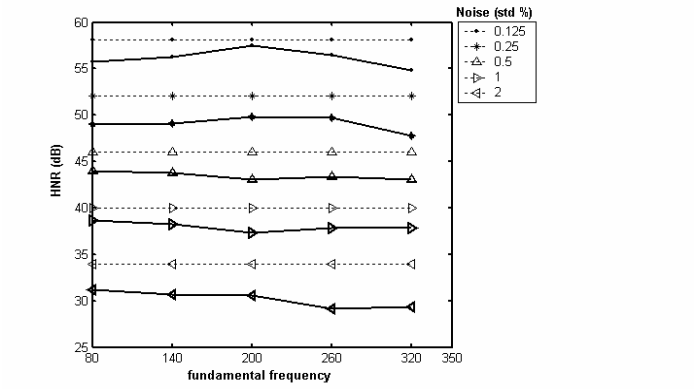

Fig.4. Estimated HNR (solid line, Qi and Hillman [2]) versus f0 for synthesized glottal source waveforms (dotted line – actual HNR).

3 Discussion

The de Krom technique [1] tends to underestimate the baseline due to the fact that minima are estimated in the baseline-fitting procedure. The Qi and Hillman approach [2] cannot match the noise level at low frequencies due to the influence of the source. Similar over- and under- estimates of the HNR for synthesized speech are also found (not shown). It was also noted that, inherent in these methods, is a poor noise floor estimate as illustrated by Figs 1 & 2 where the noise baseline estimate is shown to deviate from the actual noise level notably at low frequencies. This in effect undermines the accuracy with which the HNR is estimated by these methods.

4 Conclusions

Two existing cepstral–based HNR estimation techniques are evaluated using synthesized glottal waveforms and speech signals with *a priori* knowledge of the HNR for these signals. The methods provide reasonably consistent estimates of the HNR, however, HNR tends to be overestimated in [1] due to the baseline fitting procedure underestimating the noise levels and [2] tends to over-estimate the HNR due to the underestimate of noise levels due to the influence of the glottal source on the noise baseline. Further work will develop a technique to remove the bias due to the glottal source, thereby providing an accurate noise baseline from which to estimate the HNR.

5 References

[1] de Krom, G. “A cepstrum based technique for determining a harmonics-to-noise ratio in speech signals”. *J. Speech Hear. Res.* 36(2):254-266, 1993.

[2] Qi, Y. and Hillman, R.E. “Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals”. *J. Acoust. Soc. Amer.* 102(1):537-543, 1997.

[6] Fant, G., Liljencrants, J. and Lin, Q. G. “A four parameter model of glottal flow”, STL-QPSR 4, 1-12, 1985.