Spectral Tilt as a Perturbation-free Measurement of Noise Levels in Voice Signals

Peter J. Murphy

Department of Electronic and Computer Engineering,
University of Limerick,
Limerick, Ireland.
peter.murphy@ul.ie

Abstract

Acoustic analysis of voice quality proves useful in the objective assessment of voice disorders and for motivating new components for use in improving voice synthesis. A commonly used quantitative spectral index is the harmonics-to-noise ratio (HNR), which gives gross information regarding speech signal periodicity. However, as the measure is sensitive to all forms of waveform aperiodicities (not simply the additive random noise component of turbulent origin), it lacks specificity. Furthermore, the HNR of the radiated speech waveform has a fundamental frequency (f0) -dependence, increasing with fundamental frequency (for equal noise levels of the glottal source). Two spectral tilt measurements are applied to synthetically generated, aperiodic voice signals to investigate their sensitivity to the various forms of aperiodicity. The tilt measures are found to provide perturbation- (jitter and shimmer) free measures of noise levels in speech signals. However, for radiated speech waveforms the tilt measurements are strongly f0-dependent.

1. Introduction

Spectral measures of radiated speech and source flow waveforms prove useful for quantifying glottal characteristics (see, e.g. [1] for the former and [2] for the latter), for separating patient and normal data sets [3] and for motivating the introduction of new components for use in producing more individualistic voice synthesis [4]. The HNR is implicitly used as an indication of the ratio of the periodic component to the additive noise component in speech signals. However, many authors (see, e.g. [5],[6]) have shown that HNR is sensitive to jitter, shimmer and other waveform aperiodicites found in real voice signals. Therefore, HNR provides gross information regarding signal periodicity rather than specific information relating to an aspiration noise component. Furthermore, Murphy [7] shows that there is a fundamental frequency (f0) dependence on the HNR of the radiated speech waveform i.e. for equal values of noise at the glottal source the HNRs increase for the radiated speech waveform with increasing f0. Therefore, an appropriate comparison of glottal harmonics-to-noise ratios based on HNR measurements taken from the radiated speech waveform requires an f0 correction scheme.
An approach to provide a perturbation-free HNR measurement is given in [7]. However, the method requires highly accurate inverse filtering that retains both the glottal flow and glottal noise characteristics. The practical implementation of this method is the subject of continued research efforts. In the meantime convenient methods for extracting glottal source information/characteristics from the microphone recorded radiated speech waveform provide a pragmatic alternative.
The spectral measurements H_1-A_1 and H_1-A_3 (level difference between first harmonic and first and third formant levels, respectively) have been used to investigate first formant bandwidth and glottal adduction, respectively (see, e.g. [1]). Ideally, if spectral tilt is to be used as an indication of bandwidth/adduction across the range of f0s and aperiodicties to be found in normal and disordered phonation then the measures should be largely insensitive to variations in these latter characteristics. The present study investigates how the spectral tilt measures R_{14} and R_{24} (level difference between the energies from 0 to 1 kHz to the energy from 1 to 4 kHz and level difference between the energies from 0 to 2 kHz to the energy from 2 to 4 kHz, respectively) are affected by the aperiodicities of jitter (cyclic and random), shimmer and additive noise and by differences in f0 while formant bandwidths and glottal adduction/abduction characteristics remain constant.

2. Method

2.1. Spectral Tilt Measures (R_{14} and R_{24})

Two measures of spectral tilt are taken from the averaged modified periodograms[8]:

The ratio of the energy below 1 kHz to that above 1 kHz (R_{14}) and the ratio of the energy below 2 kHz to that above 2 kHz (R_{24}). R_{14} and R_{24} are calculated from the following equations:

$$R_{14} = \frac{\sum_{i=1}^{N/5} S_i}{\sum_{i=N/5}^{5xN/4} S_i} \qquad (1)$$

where S_i is the spectral energy at the i^{th} frequency location

$$R_{24} = \frac{\sum_{i=0}^{2xN/5} S_i}{\sum_{i=2\times N/5}^{5xN/4} S_i} \quad (2)$$

taken from the averaged modified periodogram (AMP) of a voiced speech signal, N = number of spectral estimates of the AMP covering up to 5 kHz.

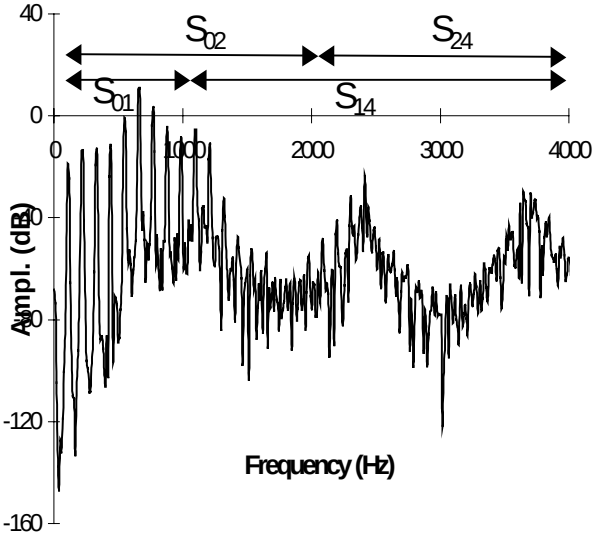


Fig.1 Averaged modified periodogram for a 110 Hz synthetically generated speech signal illustrating the signal energy ranges 0-1 kHz, 0-2kHz, 1-4kHz and 2-4 kHz used in the spectral tilt measurements.

3. Analysis

3.1. Synthetic speech signals

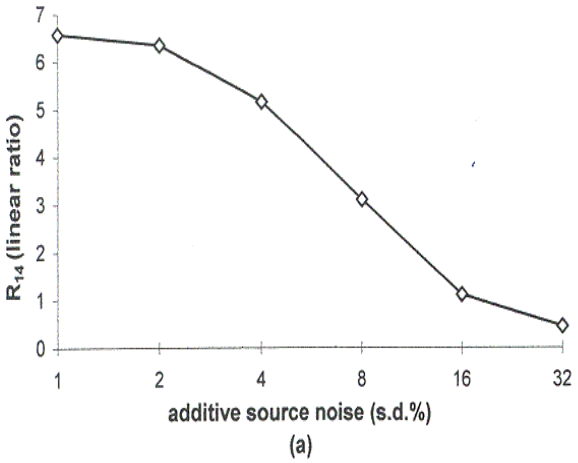
The vowel a/ is synthesized using a discrete time model for speech production with a glottal flow pulse used as the source function. Radiation at the lips is modeled by the first order difference equation $R(z)=1-z^{-1}$. Waveform aperiodicity is introduced by altering the source function. Random shimmer is introduced by adding a random variable gain factor (six doubling levels from 1 % to 32 % s.d.) to the amplitude of the pitch period impulse train prior to convolution with the glottal pulse. The pitch period is multiplied by a random number generator of given variance in order to provide the required amount of random jitter (1 % to 6 % s.d.). Random additive noise is introduced by multiplying the glottal pulse by a random noise generator arranged to give signal dependent additive noise of a user specified variance (six doubling levels from 1% to 32 % s.d.). Further signals are created for three levels of additive noise for frequencies beginning at 80 Hz and increasing in six, approximately equi-spaced steps of 60 Hz up to 350 Hz. A sampling frequency of 10 kHz is used throughout.

3.2. Analysis parameters

A Hamming window length of 2048, padded up to 4096 and hopped by 1024 points providing 8 individual spectral estimates for about 1.2 seconds of speech is used in the average modified periodograms.

4. Results

Spectral tilt measures (R_{14}, R_{24}) versus additive noise at 110 Hz are shown in Fig. 2 (a) and (b). As can be seen in the figure, as the additive noise level increases the R_{14} and R_{24} ratios decrease. R_{14} and R_{24} versus aperiodicities of random and cyclic jitter and shimmer at 110 Hz are shown in Fig. 3. Examination of these graphs reveals that the R_{14} and R_{24} ratios are largely insensitive to jitter and shimmer and even for the maximum amounts of jitter and shimmer added, the ratios are above the corresponding ratios obtained for the additive noise signals i.e. the measures provide perturbation (jitter and shimmer)-free indications of additive noise levels.



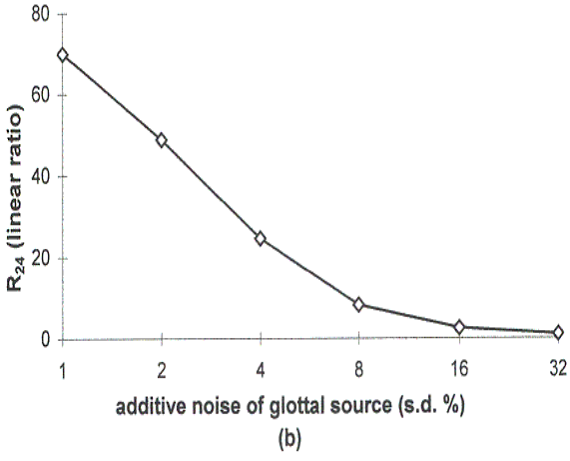


Fig.2. Spectral tilt measures R_{14} and R_{24} versus additive noise at 110 Hz are shown in (a) and (b).

The variation of R_{14} and R_{24} for three levels of additive noise at six levels of fundamental frequency from 80 Hz to 350 Hz is shown in Fig. 4., illustrating the highly non-linear variation of the indices with respect to f0 for the radiated speech waveform.

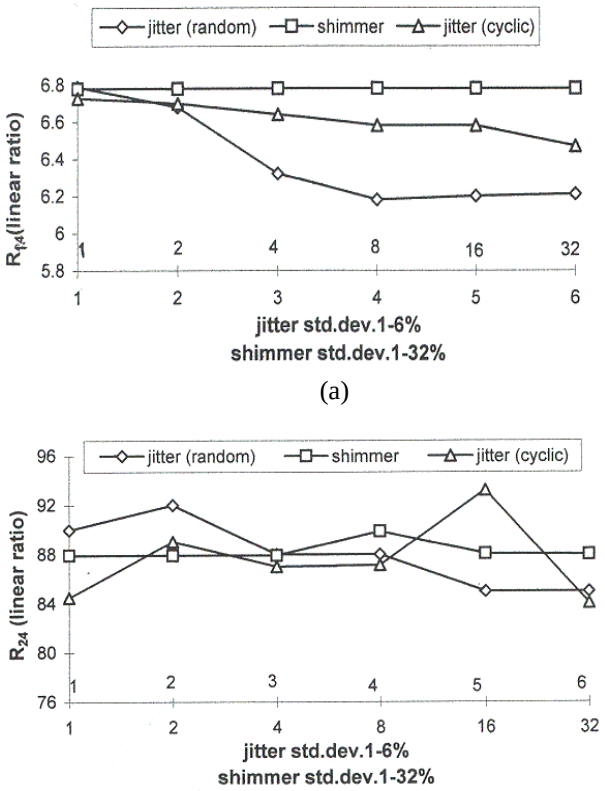


Fig.3. Spectral tilt measures (a) R_{14} and (b) R_{24} versus random jitter, cyclic jitter and shimmer at 110 Hz.

5. Discussion

Fig. 2 (a) and (b) show that measures of spectral tilt (R_{14} and R_{24}) decrease with increasing levels of additive noise. The decrease is approximately linear over a range of 30 dB. Fig. 3 (a) and (b) show that the tilt measures are largely insensitive to random jitter, cyclic jitter and shimmer perturbations. This is an interesting result because it supplies an index that is representative of the additive noise component in speech signals whilst being independent of jitter and shimmer. The result is potentially useful because the determination of a perturbation-free estimation of noise levels in voice signals is a significant problem in voice quality assessment. In addition, if jitter and shimmer are present in a speech signal, spectral tilt (and hence possibly adduction/bandwidth information) can still be estimated accurately. Again, this is noteworthy given the severe alteration in harmonic structure that is introduced as a result of jitter and shimmer [9]. However, if additive noise is present spectral tilt measures cannot be used to explicitly infer adductory and/or bandwidth information. In
relating glottal flow parameters (Fig. 5) to spectral characteristics the level of the fundamental is closely correlated with the rising portion of the flow glottogram, the rate of closure corresponds to the level of all upper partials, i.e. a higher closing speed gives rise to an increase to all higher harmonics and the spectral tilt is very dependent on the final part of the closing phase that appears after the instant of maximum airflow decrease.

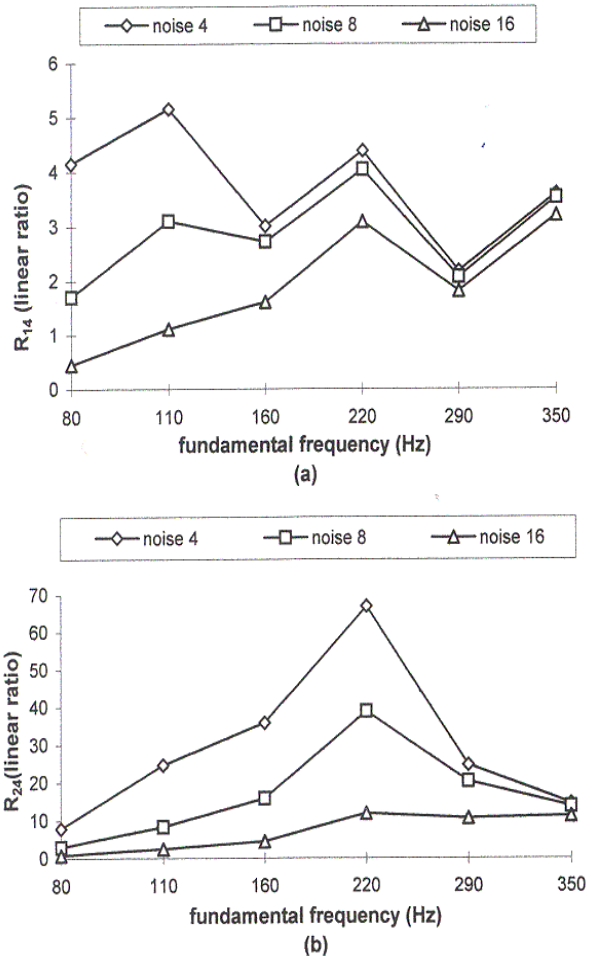


Fig. 4 (a) R_{14} and (b) R_{24} for three levels of additive noise at six levels of fundamental frequency from 80 Hz to 350 Hz.

Therefore, in order to investigate these glottal characteristics independently other spectral measurements are required. The ratio of the amplitude of the first harmonic to the amplitude of the second harmonic is thought to relate to abductory behaviour whereas the ratio of the amplitude of the first harmonic to the amplitude of the third formant is thought to relate more closely to the closing phase, though a recent study illustrates the non-unique correlation between specific time domain glottal parameters and certain spectral measurements [10].

In the present study, adductory and abductory behaviour are purposely not altered in order to firstly examine the gross effects of jitter, shimmer and additive noise on measures of spectral tilt. The present results may help to explain the apparently conflicting results regarding spectral tilt correlations with breathiness (in [1] tilt increases due to decreased abduction, in [11] tilt decreases due to greater

additive noise while in [12] tilt is not found to correlate strongly with breathiness, possibly due to a counterbalancing of the two previous effects). Therefore, to avoid ambiguity in estimating values of spectral tilt, noise sensitive and noise independent measurements are required. The tilt measure extracted from the present study is noise sensitive and has been shown to provide a perturbation- free measurement of noise levels. A method for determining noise-free tilt measurements is proposed in [7].

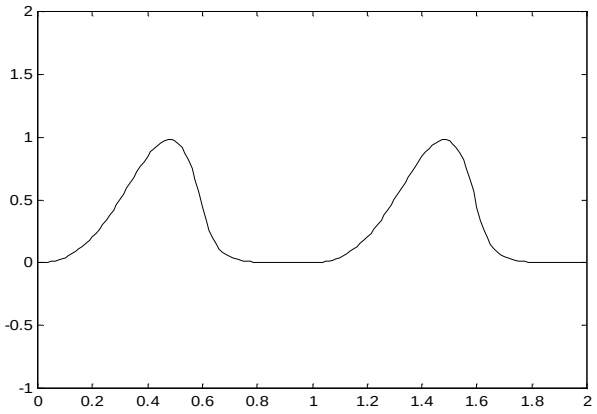


Fig.5 Two periods of a glottal flow waveform.

A practical limitation of the spectral tilt measurements is their variation with f0 (Fig. 4). The harmonics receive different resonant contributions as the fundamental frequency varies giving rise to a highly non-linear variation of tilt with fundamental frequency. In addition, the amplitude of the harmonics for the low fundamental frequency glottal signals are attenuated by approximately 2 kHz while the harmonic amplitudes for the higher fundamental frequency signals are still very prominent at 4 kHz.
In investigating the variation of the HNR and spectral tilt measures with f0 a constant filtering action has been assumed. However, the constancy of the filtering action over the entire f0 range is an oversimplification; e.g. formant frequencies of male speakers are on average 25 % lower than the formant frequencies of female speakers and many secondary filtering complications arise with breathy voice (which is a common characteristic of female speech) such as a nonlinear filtering action (open/closed glottis), the introduction of tracheal resonances and antiresonances [12], and increased formant bandwidths [1]. Nevertheless, the assumption of a constant filtering action is a more accurate approximation than scaling the filtering in accordance with the glottal flow signal.

6. Conclusion

Spectral tilt is a potentially useful candidate as a perturbation-free noise estimator in voice signals. As spectral tilt is largely insensitive to perturbations of jitter and shimmer, it should still be possible to obtain reliable abductory/bandwidth information from tilt measurements taken from signals containing such aperiodicities. Including more detailed aspects of voice source and filtering characteristics leads to a higher quality synthesis and should facilitate the future assessment of acoustic analysis techniques such as the spectral tilt.

7. References

[1] Hanson, H.M., “Glottal characteristics of female speakers; Acoustic correlates,” *J. Acoust. Soc. Amer.*, vol. 101, pp. 466-481, 1997.

[2] Alku, P. Strik, H. and Vilkman, E., “Parabolic spectral parameter - A new method for quantification of the glottal flow,” *Speech Commun.*, vol. 22, pp. 67-79, 1997.

[3] Kasuya, H. Ogawa, S., Mashima, K., and Ebihara, S., “Normalised noise energy as an acoustic measure to evaluate pathologic voice,” *J. Acoust. Soc. Amer.*, vol. 80, pp. 1329-1334, 1986.

[4] Childers, D.G., “Glottal source modeling for voice conversion,” *Speech Commun.*, vol. 16, pp. 127-138, 1995.

[5] de Krom, G., “A cepstrum based measurement of a harmonics-to-noise ratio in speech signals”, *J. Speech Hear. Res.*, vol. 36, 254-266, 1993.

[6] Ricard, G. and d’Alessandro, C., “Analysis/synthesis and modification of the speech aperiodic component”, *Speech Commun*. vol. 19, 221-224, 1996.

[7] Murphy, P.J., “Perturbation free measurement of the harmonics-to-noise ratio in voice signals using pitch-synchronous harmonic analysis,” *J. Acoust. Soc. Amer.*, vol. 105, pp. 2866-2881, 1999.

[8] Murphy, P.J. “Averaged modified periodogram analysis of aperiodic voice signals,” *Proc. Irish Signals and Systems Conf.*, Dublin, 266-271, June, 2000.

[9] Murphy, P.J., "Spectral characterization of jitter, shimmer and additive noise in synthetically generated voice signals," *J. Acoust. Soc. Amer.* vol . 107, 978-988, 2000.

[10] Swerts, M. and Veldhuis, R., “The effect of speech melody on voice quality”, *Speech Commun*. vol. 33, 297-303, 2001.

[11] Fukazawa, T., El-Assuooty, A. and Honjo, I., "A new index for the evaluation of the turbulent noise in pathological voice," *J. Acoust. Soc. Amer.* vol. 83, 1189-1192, 1988.

[12] Klatt, D.H. and Klatt, L.C., “Analysis, synthesis and perception of voice quality variations among female and male talkers,” *J. Acoust. Soc. Amer.* vol. 87, 820-857, 1990.