ANALYSIS OF SPECTRAL MEASURES FOR VOICED SPEECH WITH VARYING NOISE AND PERTUBATION LEVELS

Eoin O'Leidhin & Peter Murphy

Department of Electronic and Computer Engineering,
University of Limerick, Limerick, Ireland.
{Eoin.OLeidhin; Peter.Murphy}@ul.ie

ABSTRACT

A number of spectral measurements involving the amplitude of the first harmonic and two general spectral tilt ratios, are systematically investigated using synthetically generated speech signals with varying amounts of jitter, shimmer and noise levels. Certain spectral tilt measures are shown to provide perturbation-free measurements of noise levels in synthesized speech signals. These spectral tilt measures are then tested on real speech signals and are shown to be relatively good noise comparators in the measurement of pre- and post-treatment levels of noise in the speech of subjects with voice disorders.

1. INTRODUCTION

Extraction of glottal source characteristics has high potential for a variety of speech applications; however, deriving the total source information from the radiated speech signal remains problematic. While inverse filtering techniques are continuously being refined, there is still a need to obtain glottal source measurements directly from the speech signal and its spectrum. These measures need only the simple microphone recordings and have the potential to be easily automated. The present study examines the use of some of these measurements with regard to speech disorders and in particular in respect of their potential use for evaluating pathological voice.

As stated in [1] the main distinct sources of deviation from perfect periodicity and which in turn are the main measures that have been used to investigate the acoustic properties of pathologic voice are:

a) Variations in period of the waveform from cycle to cycle (jitter).
b) Variations in amplitude of the waveform from cycle to cycle (shimmer).
c) Noise level included in the voice signal
d) Waveshape changes
e) Non-linear phenomena such as beat frequencies – which may be caused by asynchronous coupled oscillators (i.e. asymmetrical vocal folds)

In this study, analysis of the effects of attributes (a), (b) and (c) on specific acoustic indices is presented.

Also, as stated in [2], though the HNR (harmonic-to-noise ratio) is used as an indication of the ratio between the periodic content and the noise component of the speech signal, it is in fact also sensitive to the perturbations of jitter, shimmer and other waveform aperiodicities. Therefore, HNR provides general information regarding signal periodicity rather than specific information relating to an aspiration noise component. As the extraction of a glottal signal-to-aspiration noise ratio remains an ongoing research goal, the present study also looks further into the possible use of the spectral tilt measurements R_{15} and R_{25} as perturbation-free noise indices.

It is stated in [3] that "breathy phonation is characterized by a glottal source with (1) an increased open quotient and (2) a tendency for higher harmonics to be replaced by aspiration noise". Yet there have been some conflicting findings in relation to the correlation between spectral tilt and breathiness, with Fukazawa et al. [4] stating that breathiness is associated with greater amounts of high frequency noise energy, and Hanson [5] stating that more gradual glottal closure leads to higher spectral tilt. While Hillenbrand et al. [6] and Klatt and Klatt [3] suggest that spectral tilt plays little or no role in the perception of breathy voice. In the present study a reason for these apparently conflicting results is suggested.

2. BACKGROUND

2.1. Acoustic Measurement

Some acoustic measurements extracted from the speech spectrum and averaged speech periodogram are described below:

*2.1.1. H1**

A number of studies have reported on the fact that the amplitude of the first harmonic (H1) is strongly linked to phonation type [3,6]. It is reported in [3] that out of 10 acoustic parameters only H1 and aspiration noise correlated strongly with breathiness.

As shown by Hanson [5], the harmonic values need to be adjusted (indicated by the asterisk), in order to remove the effects of the vocal tract transfer function. With regard to H1, the first formant (F1) will be the dominant vocal tract influence. This effect can be negated by subtracting the quantity $20*\log_{10}[F1^2/(F1^2-f^2)]$ from H1 where f is the frequency at which the harmonic is located.

2.1.2. H1 - H2**

The amplitude of the first harmonic (H1) relative to the amplitude of the second harmonic (H2) is used as an indicator of the relative length of the open phase to the total period of the glottal pulse, which is known as the open quotient (OQ). With regard to negating the vocal tract influences on H2 this will be carried out in a similar fashion to H1 as shown above.

2.1.3. H1 - A1*

The adjusted amplitude of the first harmonic (H1*) relative to the amplitude of the first formant peak (A1) is used to indicate the bandwidth of F1 and is also affected by source spectral tilt at lower formant frequencies. A1 is simply approximated by finding the amplitude of the strongest harmonic in the F1 region.

2.1.4. H1 - A3**

The adjusted amplitude of the first harmonic (H1*) relative to the amplitude of the strongest harmonic in the third formant region (A3) is a measure of the source spectral tilt, in particular at higher formant frequencies. Therefore the measure is strongly influenced by the rate of glottal closure. In this case as the effects the first and second formant could be considerable on the third formant region, their effects need to be removed. This is achieved by adding the quantity

$$20\log_{10}\left(\frac{\left[1-\left(\frac{F3}{F1}\right)^2\right]\left[1-\left(\frac{F3}{F2}\right)^2\right]}{\left[1-\left(\frac{F3}{\widetilde{F}1}\right)^2\right]\left[1-\left(\frac{F3}{\widetilde{F}2}\right)^2\right]}\right) \qquad (1)$$

to A3, where $\widetilde{F}1$ and $\widetilde{F}2$ are the 1st and 2nd formant frequencies of a neutral vowel. In [5] these values are given as 555 Hz and 1665 Hz respectively.

2.1.5. Spectral Tilt Measures (R_{15} and R_{25})

R_{15} and R_{25} are two spectral tilt measurements proposed by Murphy [2], which can be used to indicate the amount of noise in the speech signal but which are relatively immune to the perturbation effects of the signal. The two measures are calculated from the averaged modified periodogram [7]. The periodogram (power spectral density) is used to estimate the noise levels in voice signals and windows of the speech signal are overlapped in order to reduce the variance of the periodogram estimates. In order to reduce spectral leakage a Hamming window is used in the analysis, which gives rise to the term modified periodogram. R_{15} is the ratio of the energy below 1 kHz to the energy above 1 kHz, and R_{25} is the ratio of the energy below 2 kHz to that above 2 kHz. For this study the log of these R_{15} and R_{25} values will be used as it allows for closer comparison when varying different aspects of the speech signal.

2.2. Noise Components

A mechanical model simulation of pathological voices [8] shows that there are at least two different timing relationships between the glottal movement and the noise generation. For one noise type, the amplitude of the noise component is largest when the glottis is open, and the noise component decays when the glottis is closed. Possible reasons suggested for this type of noise were that a narrow glottal opening could cause an abnormal growth of flow velocity, or that irregular parts of the vocal folds might interfere with the airflow causing a turbulent airflow, which may lead to the noise component. For the second noise type the noise component is generated during what should be the closed phase of the glottal cycle. A probable explanation for this is that some part of the vocal folds can't maintain closure (e.g. presence of a glottal chink or the area around a nodule), leading to an airflow leakage, which generates the turbulent noise.

2.3. Perturbations effects

It is stated in [9] that for modal voice a typical jitter value would be less than 1% and a typical shimmer value is less than 0.7dB. For pathological voice these values can be considerably higher.

3. IMPLEMENTATION

The speech synthesiser used is developed in order to be able to add different types of noise and perturbations to the glottal waveform of the speech signal [10]. The synthesiser is able to add noise to the glottal flow calculated as a percentage of the glottal airflow, therefore most noise will occur at the moment of maximum glottal flow. This noise type is called multiplicative noise. Another type of noise, termed background noise, is calculated as a percentage of the average amplitude of the glottal pulse, but is otherwise independent of the glottal cycle. The final type of noise modelled is called segment noise. This is used to model noise bursts, which occur at a specific location of the glottal cycle, such as during the closing phase. The noise is added at doubling levels from 0.0625% to 8% standard deviation (s.d.). In addition, both jitter and shimmer can also be accurately controlled with the system. Jitter is varied from 0.25% to 6% s.d. while shimmer is successively doubling from 0.25% to 32% s.d. In order to obtain the acoustic measurements, 1.2 seconds of speech is synthesised. For this study the fundamental

frequency (f0) is held constant at 100 Hz. The speech signal is then widowed with a Hamming window of length 2048 and hopped by 1024 samples at a time. The final measurements are obtained by averaging the results of the windowed signals.

4. RESULTS

Table 1 shows the values for the different acoustic measures for the vowel a/ and the glottal attributes for modal speech found in [11], for different levels of multiplicative noise added to the source waveform.

s.d. %	H_1*	H_1* - H_2*	H_1* - A_1	H_1* - A_3*	R_{15}	R_{25}
0%	22.2	1.98	-11.57	-5.95	8.13	18.27
0.0625%	22.2	1.98	-11.56	-5.96	8.12	18.21
0.125%	22.2	1.98	-11.56	-6.07	8.02	18.00
0.25%	22.2	1.98	-11.58	-5.84	7.75	17.45
0.50%	22.2	1.98	-11.57	-6.41	6.75	15.13
1%	22.2	1.99	-11.59	-6.16	3.58	9.57
2%	22.2	1.98	-11.54	-7.28	-3.07	0.72
4%	21.5	2.04	-11.92	-12.91	-11.33	-8.29
8%	19.1	2.29	-13.79	-19.69	-17.71	-13.08

Table 1: Acoustic measurements for varying multiplicative noise

s.d. %	H_1*	H_1* - H_2*	H_1* - A_1	H_1* - A_3*	R_{15}	R_{25}
0%	22.2	1.98	-11.57	-5.95	8.13	18.27
0.25%	22.19	1.88	-11.47	-5.09	7.94	18.36
0.50%	22.04	1.56	-11.82	-4.57	7.26	18.86
1%	22.17	1.61	-11.23	-3.12	7.32	18.73
2%	21.79	0.81	-11.79	-4.44	7.96	19.46
3%	22.19	0.93	-9.35	-2.69	7.97	18.75
4%	21.53	0.21	-9.39	-3.93	7.74	18.61
5%	21.76	0.79	-8.91	-3.65	7.76	18.69
6%	21.69	0.47	-8.71	-5.00	7.90	18.64

Table 2: Acoustic measurements for varying jitter values.

The acoustic measurements obtained when random jitter is varied, are shown in Table 2. As can be seen, in comparison with the harmonic measurements, spectral tilt measurements (R15 and R25) are very consistent with respect to jitter variation. Meanwhile when shimmer is varied from 1% to 32% standard deviation there is found to be virtually no change in any of the measurements.

The spectral tilt measurement R25 for the various noise types and jitter and shimmer are plotted in Figure 1. Noise is varied from 0 to 8% s.d., while jitter is varied from 0 to 6% s.d., and shimmer is varied from 0 to 32% s.d. For Figure 1, this leads to different scales on the X axis, the purpose is to show that for reasonable values, noise alters the R25 value considerably while jitter and shimmer doesn't.

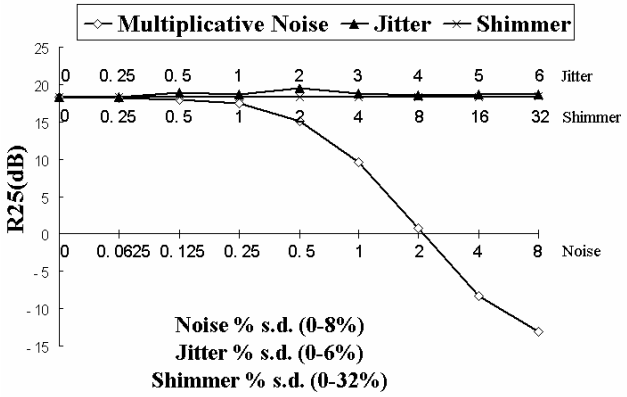

Figure 1: Spectral tilt measure R25 plotted against varying levels of noise, jitter and shimmer

It is also observed that one of the most important parameters in dictating these acoustic measurements is the speed at which the model of the glottis closes. It was noted that as the glottal closing time becomes longer, the spectral tilt becomes steeper and both R15 and R25 become larger. This is expected as with slower glottal closing times there are less high frequency components and thus the lower frequency components begin to dominate.

Finally using the data recorded from subjects with vocal disorders by Childers [11], the spectral tilt measures for vowel /IY/ were investigated comparing the measures before and after treatment. As shown in Table 3, these measures were compared with HNR values calculated using the cepstral based HNR measure.

5. DISCUSSION

As can be seen from Table 1 and Figure 1 all the spectral tilt measurements (H1*-A3*, R15 and R25) decrease considerably as the noise component of the speech signal rises. This is explained as follows; although the spectrum of the noise is approximately flat on average, the noise variance exceeds harmonic levels at the high frequencies but remains below harmonic levels at the low frequencies. This also explains the fact that the other measurements (H1*-H2*, H1*-A1) are relatively stable for all noise levels although for very high noise levels, there is some variability in the results. Also from Figure 1 it can be seen that the R15 and R25 measurements are the same for the different noise types at lower levels of noise but deviate slightly at higher noise levels.

In Figure 1 it is also shown how stable the spectral tilt measurements are for varying levels of jitter and shimmer. The reason that jitter doesn't alter the measurements R15 and R25 considerably is that although the harmonics

aren't as clearly defined as with little or no jitter, the overall shape of the spectral envelope is kept reasonably constant with varying levels of jitter. The reason that shimmer doesn't change the measurements analysed is because the spectral consequence of shimmer is to introduce sub-harmonics with amplitudes that are in direct proportion to the amplitude of neighbouring harmonics [12].

Patient (Vocal Disorder)	Stage of Treatment	HNR (dB)	R_{15}	R_{25}
Male 1 (Breathy, Hoarse)	Pre-Injection	8.52	36.59	37.01
	1 month post-injection	18.37	46.28	46.53
Male 2 (Hoarse)	Pre-Injection	23.87	41.54	42.76
	5 months post-injection	20.36	28.81	29.41
Female 1 (Breathy, weak)	Pre-Injection	16.75	37.13	38.26
	1 month post-injection	17.39	40.83	41.53
Female 2 (Vocal Fry)	Pre-Injection	14.93	37.93	38.17
	1 month post-injection	15.14	40.08	40.33
	3 month post-injection	15.4	42.55	42.64

Table 3: Comparison of spectral tilt measures to HNR values for treated subjects with various vocal disorders.

Finally the possible use of the spectral tilt measures R15 and R25 as a noise comparator is presented in Table 3. It can be seen for three of the subjects (M1, F1, F3) that the noise component in the subject's speech is reduced after treatment as is reflected in both the HNR and spectral tilt measures. In the case of Male 2 the noise has increased after treatment (possibly due to the fact that the recording was taken 5 months after the injection – and in an informal listening test it did sound worse) although once again this is seen in both the HNR and spectral tilt measures.

6. CONCLUSIONS

It was shown that varying the noise or perturbation levels causes changes in the various acoustic measures that were analysed. As shown previously [2], the spectral tilt measures R15 and R25 are insensitive to perturbation, yet reflect the noise levels reasonably well.

As actual noise estimators on their own, the R15 and R25 values would be of limited use, as the vocal tract and the glottal configuration would have a strong effect on the values. While there is the potential to use these measurements as a noise estimator in conjunction with some glottal timing index, a more likely use for these measures is perhaps in comparing the same speaker (vocal tract remains essentially the same) for improving noise levels, such as in certain pre- and post-operations (although f0 and glottal configurations may change).

7. ACKNOWLEDGEMENTS

This work is supported by Enterprise Ireland Research Innovation Fund 2002/037.

8. REFERENCES

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

[2] P.J. Murphy, "Spectral tilt as a perturbation-free estimate of noise levels in voice signals", *Proc. Eurospeech*, Aalborg, Denmark, pp. 1495-1498, 2001.

[3] D.H. Klatt, L.C. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers", *J. Acoust. Soc. Amer.*, vol.87, pp. 820-57, 1990.

[4] T. Fukazawa, A. El-Assuooty, "A new index for evaluation of the turbulent noise in pathological voice", *J. Acoust. Soc. Amer* , vol.83, pp. 1189-1193, 1988.

[5] H.M. Hanson, "Glottal characteristics of female speakers: Acoustic Correlates", *J. Acoust. Soc. Amer.,* vol.101, 466-481, 1997.

[6] J. Hillenbrand, R.A Cleveland, R.L. Erickson, "Acoustic Correlates of Breathy Vocal Quality", *J. Speech and Hearing Research*, vol.37, pp. 769-778, 1994.

[7] P.J. Murphy, "Averaged modified periodogram analysis of aperiodic signals", *Proc. Irish Signals and Systems Conf.*, Dublin, pp. 266-271, 2000.

[8] S. Imaizumi, S. Kiritani, "A Preliminary Study on the Generation of Pathological Voice Qualities", in *Vocal Physiology: Voice Production Mechanisms and Functions*, edited by O. Fujimura, Raven Press Ltd, New York, 1988.

[9] J. Hillenbrand, "Perception of aperiodicities in synthetically generated voices", *J. Acoust. Soc. Amer.,* vol.83, 2361-2371, 1988.

[10] E. O'Leidhin, P.J. Murphy, "Preliminary Glottal Source modelling for pathologic voices", *Proc. Maveba Conf.*, Florence, pp. 237-240, 2003.

[11] D.G Childers, "Speech Processing and Synthesis Toolboxes", Wiley, New York, 2000.

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