Spectra of Amplitude-Perturbed Glottal Waveforms

Jacqueline Walker and Peter Murphy

Department of Electronic and Computer Engineering
University of Limerick
Limerick, Ireland

Abstract

The spectral description of voice aperiodicities is important for the introduction of refined voice synthesis parameters and for quantifying voice disorders. A quantitative analysis of the spectral properties of glottal pulses contaminated by additive noise and shimmer is given. The results confirm contamination of higher harmonics when additive noise is present and the introduction of sub-harmonics for shimmered pulses.

1 INTRODUCTION

A study of the spectral characteristics of voice aperiodicities is useful for (i) interpreting aperiodicity/perceptual correlations (spectral correlations are expected to perform better than time-domain correlations), (ii) extracting and quantifying the degree of aperiodicity from spectral measurements (time-domain measurements are often non-independent of aperiodicity type e.g. a measurement of shimmer may be contaminated by noise) and for (iii) determining the degree to which spectral aperiodicities affect gross spectral measurements such as spectral tilt and H1-H2 ratios (amplitude of 1st and 2nd harmonics), which are intended to reflect abductory and closing phase glottal behaviour respectively. The spectral properties of voice sources containing additive noise and shimmer are presented.

2 THEORETICAL GLOTTAL PULSE MODELS

In this section, we first describe the modelling of the unperturbed glottal pulse and glottal pulse trains and provide a theoretical plot of their spectra. The predicted spectra are compared with an estimated spectrum calculated from the FFT applied to a simulated sampled glottal pulse train. Subsequent sections describe the modelling of the glottal pulse train perturbed by additive noise and by shimmer.

2.1 The Glottal Pulse Spectrum

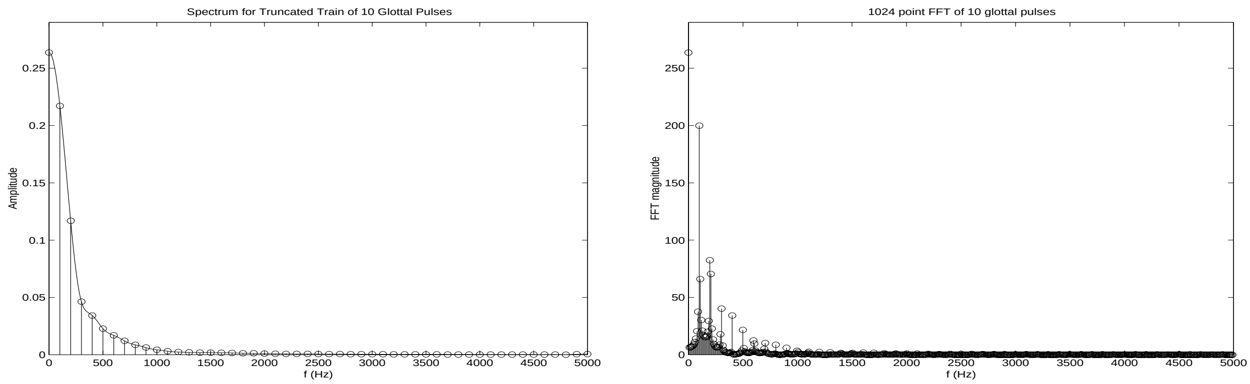
As the real glottal voice source signal is a continuous-time signal, for the theoretical analysis of the amplitude-pertubed glottal pulse, a continuous-time model may be adopted. The determination of the precise spectrum of a given glottal pulse model is possible [1] but generally involves complicated analysis. The present study describes the effects of various types of perturbations on the glottal pulse spectrum. Let us assume therefore, that for a given glottal pulse model, $g(t)$, the spectrum is given by $G(f)$. As a single glottal pulse is a time-limited signal, the spectrum will theoretically be infinite in extent, but in practice, most of the energy is concentrated at low frequencies [1].

Voiced speech, such as phonating the vowel sound /a/, involves the production of a series of glottal pulses which may be modelled by a finite glottal pulse train. Such a pulse train may be produced by windowing an infinite glottal pulse train which is itself produced by convolving the single glottal pulse with an impulse train at the desired repetition rate. The spectrum of a finite glottal pulse train is thus given by

$$G_w(f) = FG(f) \sum_{k=-\infty}^{\infty} W(f-kF) \tag{1}$$

where $W(f)$ is the spectrum of the rectangular window function. Whereas the spectrum of an infinite glottal pulse train is a line spectrum, the truncation with a rectangular window provides an interpolation by the sinc function. The result is a blurred version of the spectrum of a single glottal pulse as shown in plot (a) of Figure 1. Once again, the majority of the spectral energy is at low frequencies. For example, for a pulse train with a repetition frequency of 100 Hz, the main components are below 500 Hz.

An alternative approach to modelling the glottal pulse is to use a digital simulation using a program such as MATLAB [2]. A discrete-time determination of the spectrum of the glottal pulse is equivalent to taking a sampled version of the glottal pulse model and applying the FFT. The theoretical and simulation plots may be compared providing awareness is maintained of the effect of moving into the discrete-time domain. A sampling frequency of 10 kHz is assumed. Modelling the sampling as multiplying by an train of impulses spaced at $t_s = 0.1$ ms, the result in the frequency domain is the replication of the finite glottal pulse train spectrum at intervals of 10 kHz. As the original spectrum is theoretically infinite, aliasing will occur. The FFT of a discrete-time glottal pulse model is shown in plot (b) of Figure 1 for the glottal pulse model described by Rosenberg [3].



(a) (b)

Figure 1: (a) Theoretical spectrum of a finite, unperturbed, glottal pulse train; (b) Spectrum of discrete-time glottal pulse train model.

2.2 Effect of Additive Noise

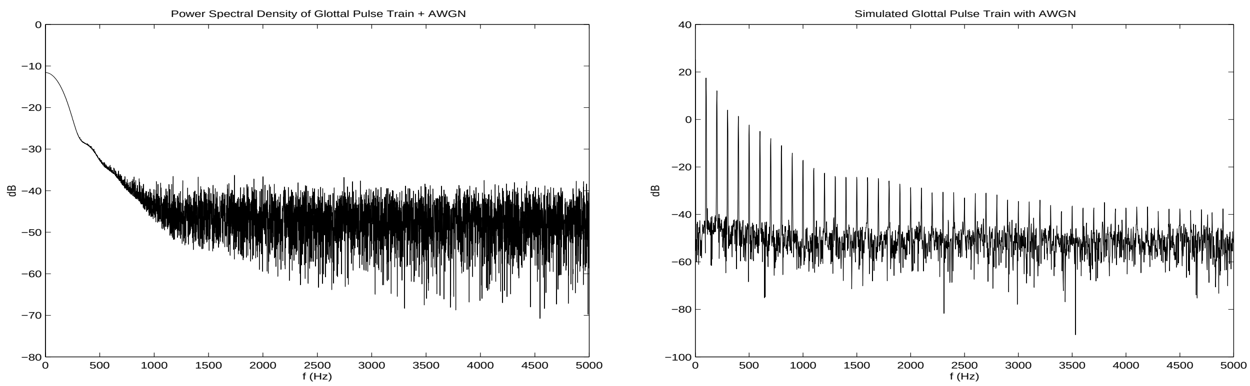
The model for additive noise is

$$g_n = g + \bar{g} + n(0, \sigma^2) \qquad (2)$$

where $n(0, \sigma^2)$ denotes additive white Gaussian noise (AWGN) with zero mean and a specified variance, σ^2, and the variance of the noise is calculated as a percentage of the mean value of the glottal pulse wavelet $\bar{g}$ [4]. Assuming that the noise is ergodic and is independent of the glottal pulse, the spectral density of the signal plus noise is

$$|G_n(f)|^2 = |G(f)|^2 + S(f) \qquad (3)$$

where $S(f)$ is the power spectral density of the noise and $|G(f)|^2$ is the power spectral density of the glottal pulse train.



(a) (b)

Figure 2: (a) Theoretical power spectral density of glottal pulse train and 2% AWGN; (b) Power spectral density of simulated glottal pulse train with 2% AWGN.

Plot (a) of Figure 2 illustrates the theoretical prediction of the effect of AWGN on the glottal pulse train spectrum. The noise spectrum was generated in MATLAB using the pseudorandom Gaussian noise generator randn() to produce zero mean AWGN with a variance of 2% of the mean value of the glottal pulse wavelet. Note that the power spectral density of the glottal pulse train is interpolated for clarity of presentation. In plot (b) of Figure 2, the estimated power spectral density of a truncated glottal pulse train lasting for 0.4s with 2% AWGN is shown. A 4096 point Hamming window was applied when taking the 4096 point Fourier transform used to calculate the estimate. The plots in Figure 2 agree with [4] in that the main effect of a low level of AWGN is to obliterate the higher frequencies of the glottal pulse spectrum. (Note that the for the noise itself, a flat spectrum is produced for all frequencies.) As a result of this effect, there may be an overestimation of the level of harmonics due to the presence of the noise at the higher frequency locations when using traditional HNR (harmonics-to-noise) estimators.

2.3 Effect of Shimmer

Shimmer is a random variation in the amplitude of the glottal pulses from period to period [4]. The kth glottal pulse will have amplitude $A(k)$ given by

$$A(k) = A + n(k) \tag{4}$$

where A is the amplitude of the glottal pulse in the absence of shimmer, and $n(k)$ is the kth sample of a Gaussian random variable with zero mean and variance which is calculated as a percentage of A. Note that the value of $n(k)$ can be either negative or positive, but the amplitude $A(k)$ is always positive. A theoretical glottal pulse train with random shimmer can be constructed by convolving a single glottal pulse with an impulse train with random amplitudes, $A(k)$.

$$g(t) \otimes \sum_{k=-\infty}^{\infty} A(k)\delta(t-kT) \tag{5}$$

The impulse train with random amplitudes is a stochastic process which can be shown to be wide sense cyclostationary with a periodic autocorrelation, $R(\tau)$ of period T [5] given by

$$R(\tau) = \sum_{j=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} E\{A(m+j)A(j)\}\delta(t-jT)\delta(\tau-mT) \tag{6}$$

where $E\{A(m+j)A(j)\}$ is the autocorrelation of the randomly shimmered amplitude weighting factors. The time-averaged autocorrelation is found by integrating over one period and, due to the presence of impulse functions in (6), is only non-zero for $j = 0$ [6].

$$\bar{R}(\tau) = \frac{1}{T} \sum_{m=-\infty}^{\infty} E\{A(m)A(0)\}\delta(\tau-mT) \tag{7}$$

As the randomly shimmered amplitude weighting factors are assumed to be generated from AWGN, their autocorrelation is [7]

$$E\{A(m)A(0)\} = \begin{cases} \mu^2, m \neq 0 \\ \sigma^2 + \mu^2, m = 0 \end{cases} \tag{8}$$

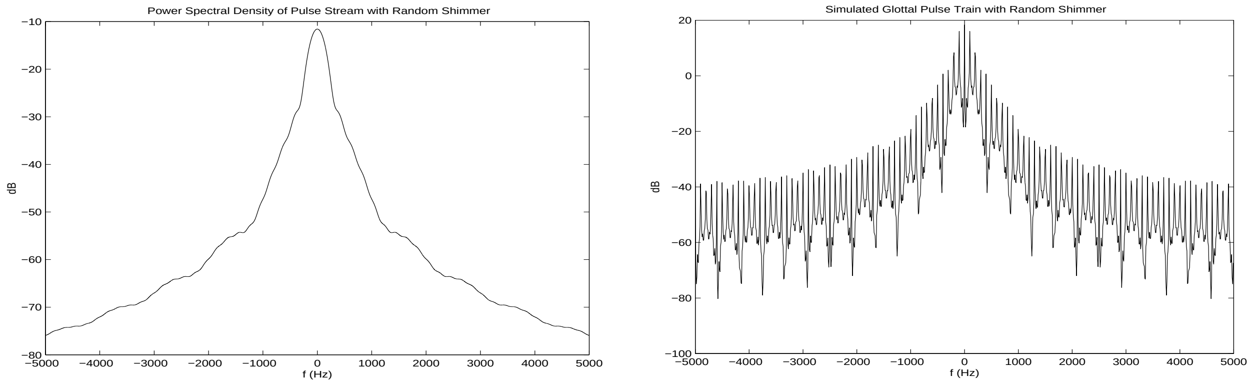
where $\mu^2 = A^2$ since the noise process is zero mean and σ^2 is that of the noise process. Hence, the power spectral density of the random impulse train is

$$\bar{S}(f) = F\sigma^2 + F^2A^2 \sum_{m=-\infty}^{\infty} \delta(f-mF) \tag{9}$$

The power spectral density of the randomly shimmered glottal pulse train is then

$$G_s(f) = F\sigma^2|G(f)|^2 + F^2A^2|G(f)|^2 \sum_{m=-\infty}^{\infty} \delta(f-mF) \tag{10}$$

Figure 3 shows the theoretical and simulated power spectral densities for glottal pulse trains with random shimmer. As before, the theoretical plot has been interpolated for clarity of comparison with the plot calculated from the simulation.



(a) (b)

Figure 3: (a) Theoretical power spectral density of glottal pulse stream with 4% random shimmer; (b) Power spectral density of glottal pulse stream (0.1s duration) with 4% random shimmer.

3 DISCUSSION

For shimmered and noise contaminated signals the $H_1 - H_2$ and $H_1 - A_1$ ratios remain unaffected so long as the unperturbed fundamental frequency is tracked in the case of shimmer and the signal harmonics are greater than the noise variance for the noise contaminated signal. When noise levels are high, extracting the Fourier coefficients (as opposed to harmonic levels) from pitch synchronous spectra and averaging the coefficients to give he $H_1 - H_2$ and $H_1 - A_1$ ratios is suggested.

4 CONCLUSION

The results presented, in combination with spectral parameterizations of glottal pulses and further examination of spectral properties of voice aperiodicities, may provide a useful scheme for modelling speech for high-quality speech synthesis. Furthermore, the spectral representations form a more appropriate basis on which to correlate perceptual evaluations of changing pulse shapes and aperiodicities, as opposed to time-domain correlations. Future work will develop these issues in greater detail

REFERENCES

1. Walker, J. “Spectra of Glottal Pulse Waveforms”, Technical Memorandum, 2000.
2. The MathWorks Inc. *MATLAB*, 1984-1999
3. Rosenberg, A. E. “Effect of glottal pulse shape on the quality of natural vowels”, *J. Acoust. Soc. Am.*, Vol. 84, pp. 583-588, 1971.
4. Murphy, P. J. “Spectral characterization of jitter, shimmer, and additive noise in synthetically generated voice signals”, *J. Acoust. Soc. Am.*, Vol. 107, pp. 978-988, 2000.
5. Papoulis, A. *Probability, Random Variables and Stochastic Processes*, 2nd ed., McGraw-Hill, 1984.
6. Howard, R. personal communication.
7. Carlson, A. B. *Communication Systems*, 3rd ed., McGraw-Hill, 1986.

ACKNOWLEDGEMENT

This work was supported in part by Enterprise Ireland Strategic Research Grant “High Quality Speech Codec for Mobile Communications”, ST/2000/100/A R/N7271.