RELATIONSHIP BETWEEN HARMONIC AMPLITUDES AND SPECTRAL ZEROS AND GLOTTAL OPEN QUOTIENT

Peter J. Murphy

Department of Electronic and Computer Engineering,
University of Limerick, Limerick, Ireland.
Peter.Murphy@ul.ie

ABSTRACT

An analysis of spectral details relating to the glottal flow waveform and its first derivative can be used to inform both formant and parametric synthesis strategies. Specifically, the current study presents a conceptual basis for the empirically known relationship between the difference in amplitude between the first and second harmonics (H1-H2) and open quotient (OQ). The position of the first spectral null and the pattern of spectral zeros are shown to contain information relevant to the duration of the open period. The analysis suggests conditions for optimum power output for specific pulse characteristics. These conditions may be important for improved naturalness of the resulting synthesized waveforms and may also be relevant to vocal performance issues.

Keywords: Glottal waveform, harmonics, spectral zeros, open quotient

1. INTRODUCTION

The need for better source modeling in formant synthesis has been highlighted for some time (c.f. [1]). Hanson and Stevens, 2002 [2] demonstrate that articulatory constraints on the voice source can inform the formant synthesis approach. It has also been shown that better source modeling can contribute to spectral modeling approaches such as the sinusoidal model and the harmonic plus noise model (HNM) [3]. The present investigation looks at how source information can be derived from the spectral details of the glottal flow.

It is a well known empirical observation that H1*-H2* (modified - to negate the influence of the first and second formant - amplitude of the first and second harmonics) relates to the open quotient (OQ) of the glottal flow [4]. Modeling studies confirm this result, but they also show that asymmetry of the waveform can also influence the measure [5]. However, it is noted that care must be taken in interpreting the model results as the underlying model variables may not behave in an unconstrained manner for real voice signals.

Flanagan, 1972 [6] presents an analytical study relating triangular waveform approximations of the glottal flow to the location and depth of spectral notches due to source zeros. The zeros are shown to depend on certain characteristics of the glottal waveform, such as symmetry. Klatt and Klatt, 1990 [1] reveal that in some cases locations of spectral zeros can be used to confirm the correct value for OQ. Earlier studies also report similar findings; [7] shows that (symmetrical and asymmetrical) triangular approximations to the glottal flow provide a good match to the zero pattern of the flow waveforms obtained from inverse filtering human voice signals. The author states that “there is a tendency in many (glottal flow spectra) for the variation to be cyclical, that is there will be a dip about every fourth or fifth harmonic.” Mathews *et al.*, 1961 [8] predict, using an s-transform analysis of the glottal pulse, that the duty factor (i.e. OQ) of the glottis determines the spacing of zeros at high frequencies. However, pitch synchronous spectra indicate zeros down to low frequencies. The authors state that glottal zeros are difficult to observe in the spectrograms of speech signals due to the pitch asynchronous nature of the processing. Carr and Trill, 1964 [9] observe a pattern of spectral zeros for one third of their data and note that for most of these patterns the zeros are multiples of the fourth or fifth harmonic. The authors also note that a zero may lie close to a formant frequency.

Ananthapadmanabha (1984) [10] notes (for modeled glottal derivative data) that the location of the first spectral null is different from the location of successive nulls and suggests that the first null location can be used as an indirect measure of the effective open time. The author also notes that the peak in the glottal derivative spectrum may not be the same as the peak due to the ‘glottal frequency’ ($2/t_p$, where t_p is the time from glottal opening to peak flow). It is noted in [11] that the return phase of the glottal derivative waveform behaves like a first order low-pass filter and that it is the main parameter for changing spectral slope. They also state that the ratio between the 'glottal frequency' and fundamental frequency, Fg/f0 varies between 0.7 to 1.6.

The current study presents a conceptual Fourier analysis of simple approximations to glottal pulses to illustrate the relationship between the spectral measures and the underlying temporal waveforms. The study presents a number of complimentary investigations; i) a conceptual interpretation of how the Fourier series coefficients are derived from glottal waveform approximations, ii) Fourier series (period, T) and Fourier transform (period T padded to 1024 points at 10 kHz sampling) estimates of synthesized half sinewave approximation waveforms, and iii) analytical representations of approximation waveforms.

2. METHOD

2.1. Synthesis data

Simple positive half-sinewave approximations to the glottal flow waveform, with OQ=0.25, eqtn. (1), OQ=0.5, eqtn. (2) and OQ=0.75, eqtn. (3) are synthesized.

$$g(t) = \sin \frac{2\pi t}{T}, 0 < t < \frac{T}{2} \qquad (1)$$
$$= 0, \frac{T}{2} \le t \le T$$

$$g(t) = \sin \frac{4\pi t}{T}, 0 < t < \frac{T}{4} \qquad (2)$$
$$= 0, \frac{T}{4} \le t \le T$$

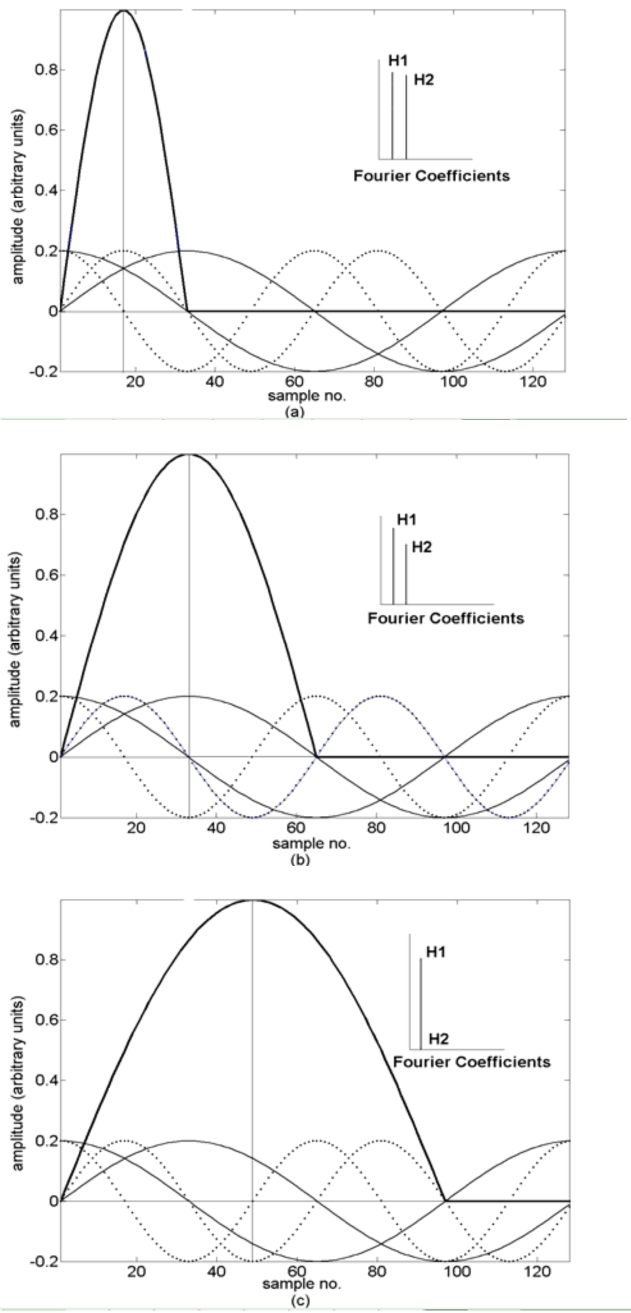
$$g(t) = \sin \frac{4\pi t}{3T}, 0 < t < \frac{3T}{4} \qquad (3)$$
$$= 0, \frac{3T}{4} \le t \le T$$

where t represents time and T represents periodic time.

2.2. Spectral characterization

A convenient procedure for postulating spectral characteristics of glottal parameters is introduced in [12]. The paper uses a graphical development that refers directly to the analytical expression for Fourier series expansion to determine spectral characteristics of jitter, shimmer and additive noise. In that study a minimum of two cycles of the waveform are required to determine spectral

Figure 1: One period of a symmetrical glottal flow waveform approximation with the sine and cosine terms for the first two Fourier coefficients a_1, b_1 (solid lines) and a_2, b_2 (dotted lines), coincident with the waveform, (a) OQ=0.25 (b) OQ=0.5 and (c) OQ=0.75.



properties of the perturbations of jitter and shimmer. In the present study, as period specific characteristics of the flow waveform are under investigation, a single glottal cycle is used in the analysis. The application of the procedure is a useful exercise because it provides a graphical description of how the Fourier series gathers its spectral estimates and hence motivates an intuitive feeling for the time/frequency relationships under investigation. The spectra for the synthesis data are then produced i.e. Fourier series coefficients are evaluated in order to test the hypotheses. In applying the Fourier series one period of a symmetrical glottal pulse with a total time record length *T* is considered, as shown in Fig. 1.

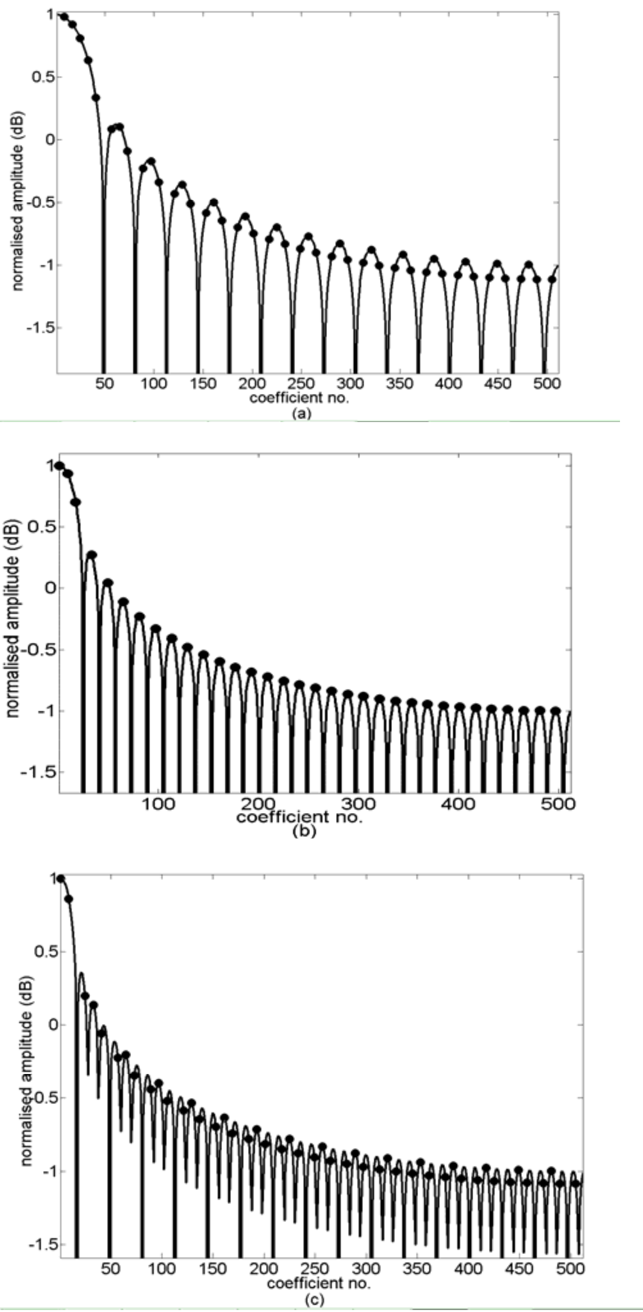
Simple approximation waveforms are examined initially as the analysis is more tractable. A positive half sinewave of OQ=0.5 is shown in Fig.1(b) with the first two Fourier coefficients superimposed. The amplitudes for the Fourier coefficients are determined by taking the product between the coefficients a_k and b_k, respectively and the underlying waveform amplitude and integrating the result. This calculation can be estimated through visually examining the figure. For coefficient a_1, it is observed that the positive half cycle matches the underlying half sinewave. Taking this product and integrating (summing) gives a value of ½. For b_1 the positive half cycle cancels the negative half cycle giving a value of zero and hence a value of ¼ for H1. For a_2 the positive and negative half cycle cancel in the product, while for b_2 the product does not cancel providing net energy for H2, though somewhat less than that for H1.

In Fig.1(a) the OQ is set to ¼. Following the same analysis as per Fig.1 it can be inferred that H2 takes on a value of ¼, while H1 is high, receiving significant contributions for both a_1 and a_2. In Fig.1(c), OQ=3/4, H1 is lower than for the previous two OQs but it still receives contributions from both a_1 and a_2. For H2 a first glottal zero is present, providing an overall large H1-H2 ratio.

The trend of H1-H2 versus OQ for this simplified analysis corresponds with what is reported in the literature and provides insight into the origin of this non-obvious relationship. The continuous and harmonic spectra for the three OQs are shown in Fig.2 (first point in each figure represents 0 Hz, 2^{nd} point is H1 etc.). The time expansion frequency compression property of the Fourier transform is observed in the figure. The H1-H2 trend is confirmed through a Discrete-time Fourier Series analysis (Table 1).

The same analysis is adopted to determine the location of zeros in the spectrum (Fig. 3). The energy contributions to the third harmonic, H3 cancel providing a first zero in the spectrum. It can be inferred from the diagram that subsequent zeros occur at higher odd harmonics.

Figure 2: Fourier series (dots) and Fourier transform estimates for one period of a symmetrical positive half sinewave glottal flow waveform approximation, (a) OQ=0.25, (b) OQ=0.5 and (c) OQ=0.75. (The Fourier transform is estimated by zero- padding the pitch period to 1024 points).



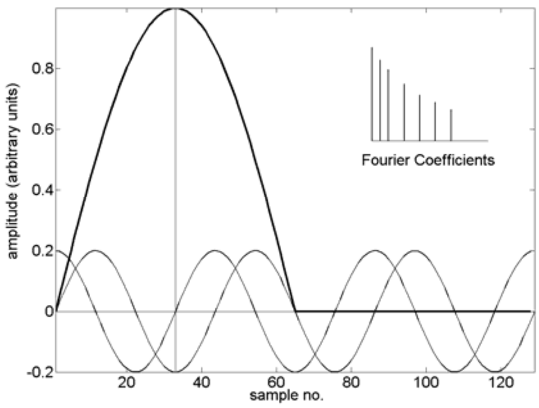
This can be confirmed analytically by taking the Fourier series of eqtn. (1) to give,

$$f(t) = \frac{1}{\pi} + \frac{1}{2}\sin\frac{2\pi t}{T} - \frac{2}{\pi}\sum_{n=2,4,6...}\frac{1}{n^2-1}\cos(\frac{n2\pi t}{T}) \quad (4)$$

Table 1: Variation of H1, H2 and H1-H2 with OQ for a Discrete-time Fourier series analysis of a symmetrical positive half-sinewave approximation to the glottal pulse.

OQ	H1	H2	H1-H2
0.25	0.98	0.93	0.05
0.5	0.98	0.7	0.28
0.75	0.89	0.0	0.89

Figure 3: One period of a symmetrical glottal flow waveform approximation (OQ=0.5) with the sine and cosine terms for the third Fourier coefficients a_3, b_3 coincident with the waveform.



3. DISCUSSION

The conceptual Fourier analysis of the approximate glottal pulses with varying OQs reveal spectral characteristics as reported in the literature regarding H1-H2 variation (c.f. [2]) and spectral zero placement (c.f. [7-9]). The Fourier series coefficients and Fourier transforms are calculated for the synthesized waveforms. The results support the graphical development. An analytical study supplies further support. Although the present study focuses on a simplified synthesis strategy the H1-H2 variation and pattern of spectral zeros are observed in both synthesized and real speech waveforms as reported in the literature.

The study highlights the potential importance of harmonic location with respect to the continuous frequency axis; this suggests conditions for maximum power output (avoidance of zeros and sampling the peaks of the continuous spectrum). It also suggests a possible mechanism for octave raising (Fig. 2(b) – higher odd harmonics coinciding with spectral zeros).

4. CONCLUSION/FUTURE WORK

The Fourier series analysis (Table 1) supports the empirical observation that as OQ increases, H1-H2 increases. The location of spectral zeros are determined by the underlying pulse shape and open period. Whether a zero occurs, depends on the sampling of the continuous frequency spectrum at harmonic frequencies. If the OQ (in the case of the symmetrical waveforms) can be represented by the division of two integers then a harmonic will fall at a zero location.

The present investigation will be extended to examine further glottal waveshape characteristics including asymmetry, and inverse filtered real data.

Once these relationships are learnt the resulting information can potentially be employed directly in HNM type synthesis systems.

5. REFERENCES

[1] Klatt, D., Klatt, L.G. 1990. Analysis, synthesis and perception of voice quality variations among female and male talkers, *J. Acoust. Soc. Am.*, 87(2):820-857.

[2] Hanson, H.M., Stevens, K.N. 2002. A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using Hlsyn, *J. Acoust. Soc. Am.*, 112(3):1158-1182.

[3] Jiang, Y., Murphy, P.J. 2002. Production based pitch modification of voiced speech, *Proceedings Intl' Conf. Spoken Lang. Processing*, Denver, Colorado, USA, Sept.: 2073-2076.

[4] Hanson, H.M. 1997. Glottal characteristics of female speakers: Acoustic correlates, *J. Acoust. Soc. Am.*, 112(3):1158-1182.

[5] Henrich, N., d'Alessandro, C., Doval, B. 2001. Spectral correlates of voice open quotient and glottal flow asymmetry: theory, limits and experimental data, *Proceedings EUROSPEECH*.

[6] Flanagan, J.L. 1972. Speech analysis, synthesis and perception, Springer Verlag.

[7] Carr, P.B., Trill, D. 1964. Long-term larynx-excitation spectra, *J. Acoust. Soc. Am.*, 36(11):2033-2940.

[8] Miller, R.L. 1959. Nature of the vocal cord wave, *J. Acoust. Soc. Am.*, 31(6):667-677.

[9] Mathews, M.V., Miller, J.E, David, Jr., E.E. 1961. Pitch synchronous analysis of voiced speech, *J. Acoust. Soc. Am.* 33(2):179-186.

[10] Ananthapadmanabha, T.V. 1984. Acoustic analysis of voice source dynamics, *STL-QPSR* 2-3, 1-24.

[11] Fant, G., Lin, Q. 1988. Frequency domain interpretation and derivation of glottal flow parameters, STL-QPSR, 2-3, 1-21.

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