ISCA Archive
http://www.isca-speech.org/archive

ITRW on Non-Linear Speech
Processing (NOLISP 03)
Le Croisic, France
May 20-23, 2003

ON THE BANDWIDTH OF A SHAPING FUNCTION MODEL OF THE PHONATORY EXCITATION SIGNAL

Jean Schoentgen
Laboratory of Experimental Phonetics
Université Libre de Bruxelles, CP 110
50, Av. F.-D. Roosevelt
B-1050 Brussels
Belgium

jschoent@ulb.ac.be

SUMMARY

The objective of the article is to present a polynomial waveshaper model of the phonatory excitation signal, as well as investigate its bandwidth when the model is driven by frequency and amplitude-modulated simple harmonic functions. The purpose of the modulation of the driving harmonics is the simulation of a time-variable intonation and accentuation, as well as phonatory regimes that are not simply periodic. The results show that the upper bound of the bandwidth of the synthetic excitation is equal to the upper bound of the bandwidth of the driving functions multiplied by the order of the model plus one. The sampling frequency must be chosen accordingly to avoid aliasing. This result can be generalized to any polynomial shaping function model.

INTRODUCTION

The phonatory excitation signal or glottal source signal is the acoustic signal that is generated at the glottis via the vibrating vocal folds and pulsatile airflow. A nonlinear shaping function model is a pair of polynomials that transform a simply harmonic input into the desired glottal excitation waveform.

We have developed a shaping function model of the glottal source signal that has the following properties (Schoentgen, 1990; Schoentgen, 2002). First, the shaping function model can be linearly fitted to observed or simulated glottal waveforms. The fit is algorithmic and constraint-free. Second, the instantaneous phonatory frequency is set by a single parameter. A change in the frequency of phonation leaves the values of the amplitude and spectral centroid of the excitation cycle unaffected. Third, the value of the spectral centroid depends on a single parameter, which leaves the phonatory frequency unaffected. Fourth, the cycle amplitude is set by a linear scaling parameter. Fifth, the model parameters fall into two categories, that is, those that fix the instantaneous values of the gain, spectral centroid and phonatory frequency on the one hand, and the polynomial coefficients that fix the default cycle shape that is typical of the speaker on the other. Sixth, the excitation cycle is zero on average and identically zero when the glottal airflow rate is a constant. The latter property is typical of acoustic signals and is added in existing models of the phonatory excitation in an ad hoc manner only (Fant et al., 1985; Klatt et al., 1990).

The mathematical properties that are summarized in the Modelling section show how a periodic glottal excitation signal may be generated via a simply harmonic driving function. The bandwidth of the periodic synthetic phonatory excitation is fixed by the order of the shaping polynomials and the frequency of the driving harmonic.

Given that the instantaneous frequency, amplitude and spectral centroid of the synthetic excitation are controlled via the linear gain of the model, as well as via the instantaneous frequency and amplitude of the driving harmonics, it is possible to synthesize glottal excitation signals whose instantaneous frequency, amplitude and spectral centroid are time-variable. The problem is that then the shaping function model must be driven by frequency and amplitude-modulated sine and cosines whose bandwidths are not zero. Consequently, the bandwidth of the synthetic output cannot be predicted anymore by means of a model that has been developed for strictly harmonic inputs only. When the signals and shaping functions have been discretized, a risk associated with driving the nonlinear model with an input that is not simply harmonic is the appearance of aliased spectral components owing to the polynomial filter. The purpose of the proposed article is therefore to study formally the bandwidth of the synthetic excitation signal when the waveshaper model is driven by an input that is not simply harmonic.

PERIODIC WAVESHAPER MODEL OF THE GLOTTAL EXCITATION SIGNAL

We have shown elsewhere that a nonlinear zero-memory model of the glottal airflow rate is the following (Schoentgen, 1990).

$$(1)\quad s(n) = G[\sum_{i=0}^{M} c_i x^i(n) + y(n)\sum_{i=0}^{M} d_i x^i(n)]$$

Model (1) is a periodic model of the glottal airflow rate that has properties (i) to (v) that are discussed in the Introduction section. Symbol *G* is the linear gain, *s(n)* is the discrete glottal airflow rate, coefficients c_i and d_i are the polynomial shaping function coefficients, and signals *x(n)* and *y(n)* are the discrete driving signals (2). Symbol *w* is the constant angular frequency multiplied by the sampling step.

$$(2)\quad x(n) = \cos(wn), \quad y(n) = \sin(wn)$$

The polynomial coefficients are obtained via constant matrix relations (3) from the Fourier series coefficients a_i and b_i of a cycle of the periodic glottal airflow rate. The values of these constant matrixes can be read out from a Pascal arithmetical triangle (Schoentgen, 1990).

$$(3)\quad \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ \ldots \end{pmatrix} = \begin{bmatrix} 1 & 0 & 2 & 0 & 6 & \ldots \\ 0 & 1 & 0 & 3 & 0 & \ldots \\ 0 & 0 & 1 & 0 & 1 & \ldots \\ 0 & 0 & 0 & 1 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 1 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix} \begin{pmatrix} c_0 \\ c_1/2 \\ c_2/4 \\ c_3/8 \\ c_4/16 \\ \ldots \end{pmatrix}, \quad \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ \ldots \end{pmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 & 2 & \ldots \\ 0 & 1 & 0 & 2 & 0 & \ldots \\ 0 & 0 & 1 & 0 & 3 & \ldots \\ 0 & 0 & 0 & 1 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 1 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix} \begin{pmatrix} d_0/2 \\ d_1/4 \\ d_2/8 \\ d_3/16 \\ d_4/32 \\ \ldots \end{pmatrix}.$$

More recently, we have shown that a shaping function model of the phonatory excitation signal can be obtained by taking the derivative with respect to the phase of the driving signals *x(n)* and *y(n)* in model (1) (Schoentgen, 2002*)*. Model (1) then turns into model (4), which is a model of the glottal excitation signal *e(n)*. Model (4) has the required pulse shape, is zero on average, and identically zero when the glottal airflow rate is a constant.

$$(4)\quad e(n) = G[-y(n)\sum_{i=1}^{M} c_i i x^{i-1}(n) + \sum_{i=0}^{M} d_i x^{i+1}(n) - y^2(n)\sum_{i=1}^{M} d_i i x^{i-1}(n)]$$

Matrix relations (3) and model (4) show that the bandwidth of the output of the waveshaper is equal to *M+1* multiplied by the frequency of the driving cosine, because *M+1* is the exponent of the monomial of the highest power. Indeed, model (4) generates the required frequency components via the products and powers of the harmonic driving functions only.

APERIODIC WAVESHAPER MODEL OF THE GLOTTAL EXCITATION SIGNAL

When model (4) is used for synthesis purposes, then it may be multiplied by a linear gain *G* that is not constant, and driven by functions *x(n)* and *y(n)* whose instantaneous amplitudes and frequencies are time-variable. Symbol *w(n)* is the discrete instantaneous angular velocity, that is, the instantaneous continuous angular velocity multiplied by the sampling step.

$$(5)\quad x(n) = A(n)\cos[q(n)], \quad y(n) = A(n)\sin[q(n)]$$
$$\textit{with}\ q(n) = \sum_n w(n)$$

A formal problem, which may have practical consequences, is that now excitation signals (5) are modulated harmonics with non-zero bandwidths. Consequently, the bandwidth of the output of the waveshaper is not anymore an integer multiple of the frequency of the driving cosine.

The problem of determining the bandwidth of the model output may be addressed as follows. We assume that the amplitude or frequency of the driving cosine evolves over an interval of length L (in number of samples). We then extend the modulated cosine over the interval $(-L,L)$ as an even function so that its Fourier series comprises cosines only. Furthermore, we assume that the bandwidth of the driving function is fixed so as to comply with the sampling theorem. The driving function may then be rewritten as follows. Coefficients A_k are the coefficients of the Fourier sum, and index K is its order.

$$(6)\quad x(n) = \sum_{k=0}^{K} A_k \cos\left(\frac{kpn}{L}\right)$$

Because in model (4) the exponent of the monomial of the highest power is equal to $M+1$, we consider hereafter Fourier sum (6) elevated to that power. The reason is that in a polynomial waveshaper model, the monomial with the highest power also generates the spectral components of the highest frequency.

The multinomial theorem gives the following identity, where the sum on the right-hand side is over all non-negative integers summing to $M+1$.

$$(7)\quad \left[\sum_{k=0}^{K} A_k \cos\left(\frac{kpn}{L}\right)\right]^{M+1} = \sum \frac{(M+1)!}{n_0!\,n_1!\,\ldots n_K!} A_0^{n_0} A_1^{n_1} \cos^{n_1}\left(\frac{pn}{L}\right) \ldots A_K^{n_K} \cos^{n_K}\left(\frac{Kpn}{L}\right)$$

Generally speaking, the higher the power of a cosine, the higher the frequencies of its spectral components are. This can be shown formally be means of Euler’s formulae.

$$Z = \cos q + j\sin q$$
$$(Z + Z^{-1})^n = 2^n \cos^n q$$
$$Z^n = \cos nq + j\sin nq$$
$$Z^{-n} = \cos nq - j\sin nq$$
$$Z^n + Z^{-n} = 2\cos nq$$

Consequently, the spectral components with the highest frequency are due to the term in sum (7) for which $n_K = M+1$ and $n_k = 0$ otherwise. By substitution and by making use of the binomial theorem we may accordingly expand this term as follows.

$$(8)\quad \cos^{M+1}\left(\frac{Kpn}{L}\right) = \frac{1}{2^{M+1}}\left\{\binom{M+1}{0}\cos\left[\frac{(M+1)Kpn}{L}\right] + \binom{M+1}{1}\cos\left[\frac{(M-1)Kpn}{L}\right] + \right.$$
$$\binom{M+1}{2}\cos\left[\frac{(M-3)Kpn}{L}\right] + \binom{M+1}{3}\cos\left[\frac{(M-5)Kp}{L} + \ldots\right.$$

That is, the frequency of the spectral component with the highest frequency is equal to $\frac{(M+1)Kpn}{L}$. In other words, the upper bound of the bandwidth of the output of the polynomial shaping function model (4) is equal to *(M+1)* multiplied by the upper bound of the bandwidth of the driving signal (6).

To avoid aliasing, the upper bound of the bandwidth of the model output must be less than half the sampling frequency, that is,

$$\frac{(M+1)K}{L} < 1.$$

CONCLUSION

When a polynomial shaping function model of the phonatory excitation signal is driven by frequency or amplitude-modulated harmonics, the upper bound of the bandwidth of the synthetic excitation is equal to the upper bound of the bandwidth of the driving functions multiplied by the order of the model plus one. This result can be generalized to any polynomial shaping function model. The sampling frequency must be chosen accordingly to avoid aliasing.

REFERENCES

Fant, G., Liljencrants, J., Lin, Q. (1985), "A four-parameter model of glottal flow", Speech Transmiss. Lab.-Quart. Progr. Stat. Rep. 4, 1-13.

Klatt, D., Klatt, L. (1990), "Analysis, synthesis, and perception of voice quality variations among female and male talkers", J. Acoust. Soc. Am. 87, 820-857.

Schoentgen, J. (1990), "Nonlinear signal representation and its application to the modeling of the glottal waveform", Speech Comm. 9, 189-201.

Schoentgen, J. (2002), “Analysis and synthesis of the phonatory excitation signal by means of a pair of polynomial shaping functions”, Proceedings International Conference on Spoken Language Processing, Denver, 981-984.