SPECTRAL CORRELATES OF GLOTTAL WAVEFORM MODELS: AN ANALYTIC STUDY

Boris Doval *Christophe d'Alessandro*
LIMSI-CNRS, BP 133, F91403 Orsay, France.
E-mail: doval limsi.fr cda limsi.fr

ABSTRACT

This paper deals with spectral representation of the glottal flow. The LF and the KLGLOTT88 models of the glottal flow are studied. In a first part, we compute analytically the spectrum of the LF-model. Then, formulas are given for computing spectral tilt and amplitudes of the first harmonics as functions of the LF-model parameters. In a second part we consider the spectrum of the KLGLOTT88 model. It is shown that this model can be modeled in the spectral domain by an all-pole third-order linear filter. Moreover, the anticausal impulse response of this filter is a good approximation of the glottal flow model. Parameter estimation seems easier in the spectral domain. Therefore our results can be used for modification of the (hidden) glottal flow characteristic of natural speech signals, by processing directly the spectrum, without needing time-domain parameter estimation.

1. INTRODUCTION

Accurate processing of the vocal flow characteristics is needed for dealing with voice quality in high quality speech synthesis. In the context of synthesis, a frequency-domain approach appears desirable, because voice quality is better described by spectral parameters. In [7] and [1], the main spectral parameters found for synthesizing voices with different qualities are: 1/ spectral tilt; 2/ amplitude of the first few harmonics; 3/ increase of first formant bandwidth; 4/ noise in the voice source (these latter parameters will not be considered here, see for instance [7] [5]). However, recent work on glottal flow parameterization concerned mostly time-domain models, and no analytical formulas are available for description of the spectral behaviour of such models. In all the studies we have been able to locate (for instance [4], [6], [7]) the spectrum is obtained by Fourier transform of the glottal waveform. Therefore, little insight is brought on the role played by each individual component of the waveform in the spectral domain, no analytic formulas are provided for the spectrum, and no spectral model of the glottal flow are proposed.

In this paper, we study the spectral correlates of two classical glottal flow models (the LF-model presented in [3], and the KLGLOTT88 presented in [1]). In a first part, we compute analytically the spectrum of the LF-model. Then, formulas are given for computing spectral tilt and amplitudes of the first harmonics as functions of the LF-model parameters. In a second part we consider the spectrum of the KLGLOTT88 model. Both the magnitude and the phase spectra are studied. It is shown that this model can be represented in the spectral domain by an anticausal linear filter with a double real pole and a simple real pole. A direct correspondance is found between time-domain and spectral parameters.

2. LF-MODEL

2.1. Parameters of the model

The LF-model is a five parameter model of the glottal flow derivative $U_g'(t)$ [3]. The five parameters commonly used to describe the LF-model are : T_0, E_e, R_g, R_k, R_a. T_0 is the fundamental period; it will only change the harmonic frequencies. E_e is the maximum flow declination rate; it will only change the overall harmonic amplitudes. R_g is the ratio of T_0 over twice the peak flow time T_e (see figure 1). It behaves much like the open quotient O_q. The spectral effect of an increased R_g is to expand the frequency scale, resulting in shifting energy from low frequency harmonics to medium frequency harmonics. R_k is the inverse of the speed quotient : $R_k = (T_e - T_p)/T_p$; it will change the waveform skewness, and will essentially affect the first harmonics amplitude. R_a mesures the duration of the return phase : $R_a = T_a/T_0$; it will change the spectral tilt adding a $-6dB/oct$ above a frequency which depends on R_a, R_g and R_k, and then will essentially affect high order harmonics amplitude. The open quotient is related to both R_g and R_k : $O_q = (1+R_k)/(2R_g)$. See [3], [4] and [2] for details on LF-model parameters. The LF-model can produce a great variety of waveforms with the different parameter settings. But a given set of parameters does not ensure to give a plausible speech waveform. In order to do so, the parameters must satisfy their theoretical ranges : $E_e > 0$, $T_0 > 0$, $R_g > 0.5$, $1 > R_k > 0$, $R_a > 0$. But they must also verify the following equations : $R_k < 2R_g - 1$, which ensures that the closing time is inside the period, and $R_a < 1 - (1 + R_k)/(2R_g)$, which ensures that the return phase is a decreasing exponential. Furthermore, if $R_k > 0.5$ then the negative maximum of the flow derivative is no longer E_e. Thus, to keep the meaning of E_e as the maximum flow declination rate, one must force $R_k < 0.5$.

2.2. Spectrum of the LF-model

We computed the derivative spectrum of the LF-model. The result is given in Equation (1), where the variables E_0, ω_g, T_e, T_a, and ϵ are functions of the model parameters, and where a is obtained solving an implicit equation. The reason for an implicit equation is the condition of zero net gain of flow during a fundamental period which implies area balance in the flow derivative.

$$\tilde{U_g'}(\nu) = E_0 \frac{1}{(a - j2\pi\nu)^2 + \omega_g^2} \times [\omega_g + \tag{1}$$
$$\exp((a - j2\pi\nu)T_e)((a - j2\pi\nu)sin(\omega_g T_e) - \omega_g cos(\omega_g T_e))]$$
$$+ E_e \frac{\exp(-j2\pi\nu T_e)}{\epsilon T_a j2\pi\nu(\epsilon + j2\pi\nu)} \times$$
$$[\epsilon(1 - \epsilon T_a)(1 - \exp(-j2\pi\nu(T_0 - T_e)) - \epsilon T_a j2\pi\nu]$$

2.3. Spectral correlates of LF-model parameters

With the help of the analytic expression of the LF-model spectrum, one can obtain the following results on the spectral correlates of the LF-model:

2.3.1. *Spectral tilt*

The spectral tilt is an important parameter of voice quality, especially for female voices [6]. It is related to the spectrum behaviour when the frequency tends towards $+\infty$. If the parameter R_a is set to 0, then $|\tilde{U}'_g(\nu)| \sim E_e/(2\pi\nu)$ when $\nu \rightarrow +\infty$, which corresponds to a spectral slope of $-6dB/oct$. This is a $3dB$ approximation of the spectrum above frequency $\max(2a, \frac{4\pi R_g}{T0}\cot(\pi(1+R_k)))$, which is given by the second prominent term in the development. If R_a is not equal to zero, then an extra $-6dB/oct$ is added to the spectrum, leading to a $-12dB/oct$ spectral slope, above a cutoff frequency which can be computed as $f_c = F_a + a/2\pi + \frac{R_g}{T0}\cot(\pi(1+R_k))$, where $F_a = 1/2\pi R_a$. In comparison to the predicted cutoff frequency value of F_a given by Fant [1] [3], this analytically calculated value gives a correction term that is not negligible : for instance, with $R_g = 1.3$, $R_k = 0.3$ and $R_a = 0.1$, (which corresponds to a plausible speech waveform, see figure 1) then $F_a = 160Hz$ although the cutoff frequency is equal to $f_c = 290Hz$; in this case, taking F_a instead of f_c leads to a more than $5dB$ error in the determination of the spectral tilt. Figure 2 illustrates the effect of the parameter R_a. Notice that the amplitude of the first harmonics is also affected by this parameter. In conclusion, the spectral tilt depends mostly on the parameter R_a. This parameter is responsible for an extra $-6dB/oct$ attenuation above frequency f_c. However, f_c depends also on R_g and R_k according to the analytic expression $f_c = F_a + a/2\pi + \frac{R_g}{T0}\cot(\pi(1+R_k))$. Thus, contrary to [1] [3], f_c cannot always be approximated by F_a.

2.3.2. *First harmonics*

In a similar manner, one can study the low frequency harmonic amplitudes. Of particular interest is the ratio $H1 - H2$, where $H1$ and $H2$ are the amplitudes of the first two harmonics (in dB). Figure 2 shows the variation of this ratio as a function of R_k. As can be seen, $H1 - H2$ has a range of about $10dB$ for common parameter ranges $0.3 < R_k < 0.6$ and $1.0 < R_g < 1.3$. The amplitude ratio of the two first harmonics depends mostly on the open quotient and the speed quotient (or equivalently to R_g and R_k). Figure 2 illustrates the effect of R_k on the amplitude of the first harmonics. Changes in spectral tilt are also noticeable. This ratio increases with the open quotient and its range increases with R_k as shows the $1dB$ approximation : $H1 - H2 = 12(\frac{O_q}{0.7})^2(1 - (1 - \frac{R_k}{0.7})^2) - 6$.

3. KLGLOTT88 MODEL

3.1. Parameters of the model

The KLGLOTT88 model is a four-parameters model of the glottal flow [1]. The model is split in two parts: first a three-parameters time-domain description of the flow $U_k(t)$, and second a one-parameter spectral tilt filter F_k. The time-domain parameters are T_0, the fundamental period, O_q the open quotient, and AV the amplitude of voicing, which gives the peak flow (the maximum value of $U_k(t)$ being AVT_0). The spectral tilt filter F_t is a first order filter with one parameter, the cutoff frequency f_t at which an additional -6 dB/oct slope is imposed to the glottal flow spectrum. In this model, O_q is used for modeling the low-frequency part of the spectrum, and f_t is used for modeling the behaviour of higher harmonics. $U_k(t)$ is described in time-domain by a third degree polynomial with two parameters $U_k(t) = at^2 - bt^3$, with $a = (27AV)/(4T_o O_q^2)$ and $b = (27AV)/(4T_0^2 O_q^3)$.

3.2. Spectrum of the KLGLOTT88

The spectrum of the KLGLOTT88 model is given by:

$$\tilde{U}_k(\nu) = \frac{27jAV}{2O_q(2\pi\nu)^2}\left[\frac{j\exp(-j2\pi\nu O_q T_0)}{2} + \frac{1+2\exp(-j2\pi\nu O_q T_0)}{2\pi\nu O_q T_0} + 3j\frac{1-\exp(-j2\pi\nu O_q T_0)}{(2\pi\nu O_q T_0)^2}\right] \quad (2)$$

This equation shows that the parameter $av = AV/O_q$ is important for studying the spectral behaviour of the KLGLOTT88 model. When ν tends to infinity, Equation (2) is equivalent to its first term:

$$\tilde{U}_k(\nu) \sim \frac{-27av\exp(-j2\pi\nu O_q T_0)}{4(2\pi\nu)^2} \quad (3)$$

Therefore, the spectral slope is -12 dB/oct. The principal term of the series expansion of Equation (2) when ν tends to 0 is given by:

$$\tilde{U}_k(\nu) \rightarrow \frac{9}{16}av(O_q T_0)^2 \quad (4)$$

and the spectral slope of $\tilde{U}_k(\nu)$ is null in 0. The magnitude spectrum of $\tilde{U}_k(\nu)$ has a -12 dB/oct slope for high fequencies, and is constant for low frequencies. It can be represented by a second-order low-pass filter. The cutoff frequency of this filter can be computed as the crossing point of the lines defined by Equations (3) (4). This cutoff frequency f_k is given by:

$$f_k = \frac{\sqrt{3}}{\pi}\frac{1}{O_q T_0} \quad (5)$$

f_k depends only on O_q and T_0. Therefore, one can find a very simple spectral interpretation of the time-domain parameter O_q: it defines the cutoff frequency of a second order low-pass filter.

3.2.1. *Spectral tilt*

The waveform $U_k(t)$ is filtered by a first order low-pass filter. This filter is defined by its cutoff frequency f_t, and add an extra -6 dB/oct attenuation. Therefore, the magnitude spectrum of the KLGLOTT88 model can be split in three regions. In a first region, between frequencies 0 and f_k the spectral slope is 0 dB/oct, and the constant value is given by Equation (4). In a second region, between f_k and f_t, the spectral slope is -12 dB/oct. In a third region, above f_t, the spectral slope is -18 dB/oct. It must be pointed out that this representation is very interesting, because there is a one-to-one correspondance between the breakpoints in the magnitude spectrum and the independant time-domain parameters of the model. This approximation is rather good, as it is illustrated in Figure 3, which compares the magnitude spectrum of the KLGLOTT model and the magnitude spectrum predicted by the third-order low-pass filter model.

3.2.2. *Phase spectrum*

The analytic formula for the complex spectrum in Equation (2) is also useful for studying the phase spectrum, and thus the time-domain glottal pulse shape. The phase of $\tilde{U}_k(\nu)$ computed with Equation (2) can be split in a linear component and a non-linear component. The linear component is only due to the delay between 0 and the epoch of glottal closure. If this component is removed, only a non-linear component remains. This component is linked to the glottal flow shape. Figure 4 displays this component for various values of O_q. Such a phase spectrum is close to the phase gain of a low pass filter, except that it corresponds to a anticausal impulse response. As a matter of fact, removing the linear phase component is equivalent to shifting the whole waveform in time domain. Then the waveform becomes anticausal.

Figure 4 compares the phase spectrum of $\tilde{U}_k(\nu)$ and the phase spectrum of the second-order low-pass filter, estimated using the magnitude spectrum. The two signals are close together. However, they are different in nature: the KLGLOTT88 model can be interpreted as a finite impulse response system, when the filter is an infinite impulse response system. The phase of the KLGLOTT88 model shows some oscillations that are linked to the finite duration of the impulse reponse of the correponding system. Truncation of the infinite impulse response of the corresponding filter to $O_q T_0$ will introduce similar oscillation.

3.3. Spectral modeling of the KLGLOTT88 model

In summary, we found that, using the analytic formulation of the spectrum, is is possible to design an all-pole filter which is comparable to the KLGLOTT88 model. This filter is a 3 order low-pass filter, with a double real pole, and a simple real pole. The simple real pole is given directly by the f_t parameter of the KLGLOTT88 model. The double real pole is linked directly to the open quotient O_q of the KLGLOTT88 model. It is also possible to model the low-frequency region using a pair of complex-conjugate poles instead of a double real pole. This might be better for representing the "glottal formant" that can be found sometimes in actual speech. If one wants to preserve the glottal pulse shape it is necessary to design an anticausal filter. Figure 5 give an example of an anticausal impulse response derived from the KLGLOTT88 model parameters. It is compared to the KLGLOTT88 model in time domain. If one wants to preserve the finite duration property of the glottal pulse, it is necessary to truncate the impulse response of the filter.

4. SUMMARY AND CONCLUSION

In this paper we show that it is possible to model in the spectral domain accurate description of the glottal flow characteristics. Contrary to previous work, all our results are analytic. It is possible to switch from frequency to time domain or from time to frequency domain with the help of exact formulas. This formulation allows for spectral modeling. We show that simple spectral correlates of the time-domain features of the LF and KLGLOTT88 glottal flow models can be derived. In addition, in the case of the KLGLOTT88 model, a one-to-one correspondance between time-domain parameters and spectral parameters is given, for a very simple third order filter defined with only two different poles, and an anticausal impulse response.

These results are challenging the more traditional time-domain approaches of glottal modeling. Spectral modeling has a number of advantages. Time domain parameters estimation is difficult, because is depends on inverse filtering and non-linear estimation techniques. On the contrary, all-pole parameters estimation is well understood. It opens a new way for glottal parameters estimation in speech. Another application of this study is voice quality modification in the frequency domain : the modification of any time-domain parameter can be simulated directly on the speech spectrum by using the analytical formulas, and without needing time-domain parameter estimation nor inverse filtering.

REFERENCES

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

[2] Fant G., Kruckenberg A., Liljencrants J., and Bavegard M. Voice source parameters in continuous speech. transformation of lf-parameters. In *ICSLP*, pages 1451–1454, Yokohama, 1994.

[3] Fant G., Liljencrants J., and Lin Q. A four-parameter model of glottal flow. *STL-QPSR*, 85(2):1–13, 1985.

[4] Fant G. and Lin Q. Frequency domain interpretation and derivation of glottal flow parameters. *STL-QPSR*, 88(2-3):1–21, 1988.

[5] Richard G. and d'Alessandro C. Analysis/synthesis and modification of the speech aperiodic component. *Speech Communication*, 19:221–244, 1996.

[6] Gauffin J. and Sundberg J. Spectral correlates of glottal voice source waveform characteristics. *Journal of Speech and Hearing Research*, 32:556–565, 1989.

[7] Hanson H. M. Glottal characteristics of female speakers. In *PhD Thesis*. Harvard Univ., 1995.

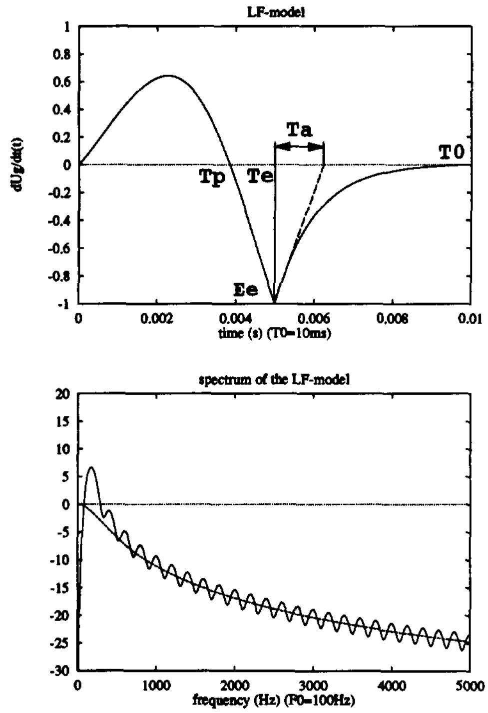

Figure 1. LF-model flow derivative (top) ($R_g = 1.3$, $R_k = 0.3$ and $R_a = 0.1$) and flow second derivative spectrum (bottom) computed using the analytic expression of $\tilde{U}_g'(\nu)$. Superimposed to the spectrum is the equivalent first order low-pass function of cutoff frequency f_c.

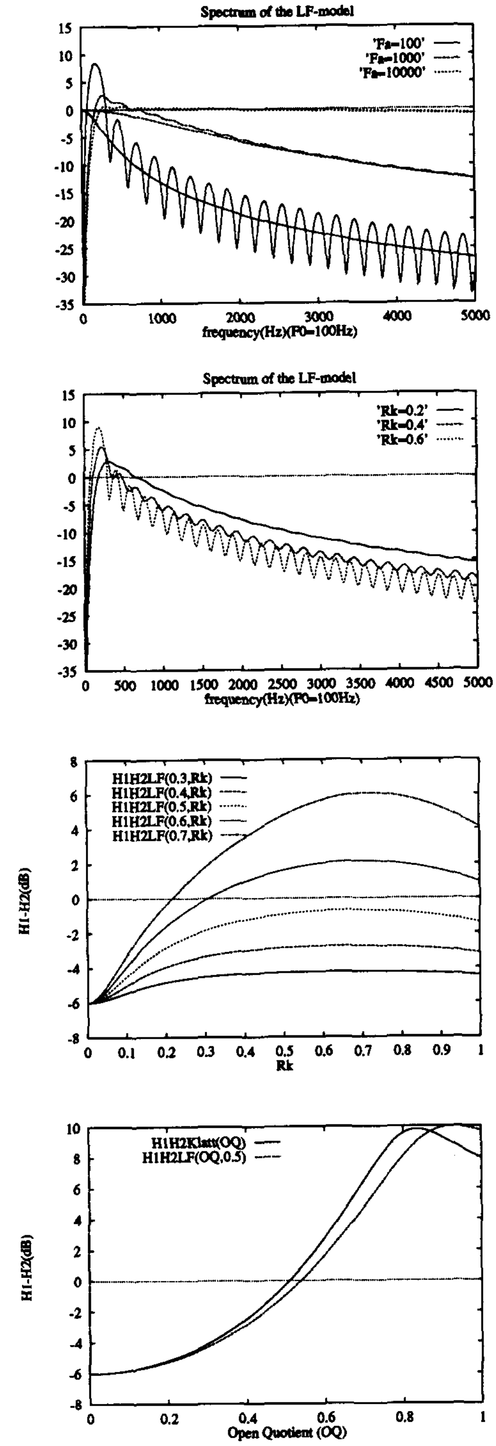

Figure 2. Top: effect of $F_a = 100, 1000, 10000Hz$ ($R_g = 1.5$, $R_k = 0.3$). Superimposed to the flow second derivative spectrum is the equivalent first order low-pass function of cutoff frequency f_c. Second from top: effect of $R_k = 0.2, 0.4, 0.6$ ($R_g = 1.5$, $F_a = 500$). Third from top: Amplitude ratio of the two first harmonics for LF-model in function of R_k for different values of O_q (from 0.3 to 0.7). Here $R_a = 0$. Bottom: same ratio for LF-model and KLGLOTT88-model in function of O_q. The intrinsic speed quotient for KLGLOTT88 corresponds to $R_k = 0.5$.

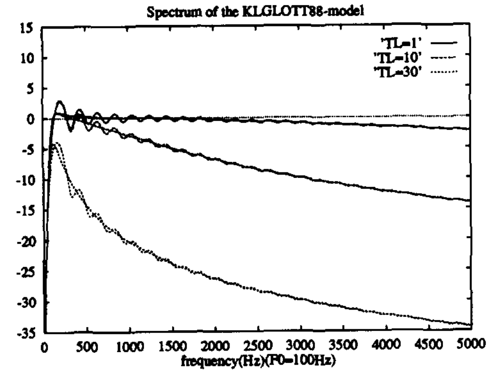

Figure 3. Magnitude spectrum of the flow second derivative of the KLGLOTT88 model, for $O_q = 0.5$, and various values of f_t (corresponding to an attenuation of TL dB at 3 kHz) . The magnitude spectrum predicted by a filter model is superimposed.

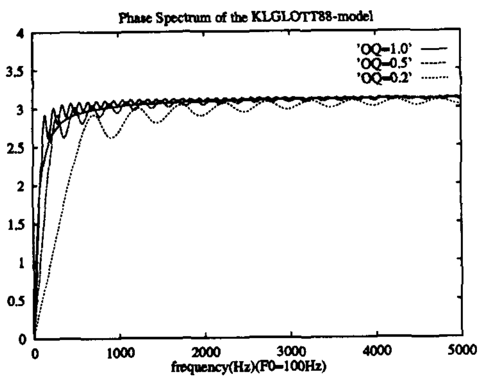

Figure 4. Phase spectrum of the KLGLOTT88 model $U_k(t)$, for various values of O_q. The phase spectrum predicted by an anticausal filter model is also plotted. TL=0

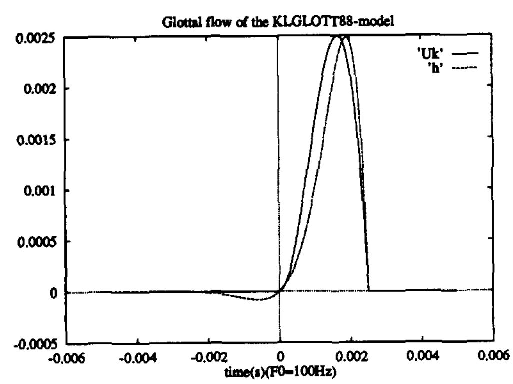

Figure 5. Comparison of the KLGLOTT88 model $U_k(t)$, and the impulse response of the corresponding anticausal filter.