Unified Analysis of Glottal Source Spectrum

Ixone Arroabarren, Alfonso Carlosena

Universidad Publica de Navarra
Dept. Electrical and Electronic Engineering
Campus de Arrosadia, E-31006 Pamplona, Spain

ixone.arroabarren@unavarra.es, carlosen@unavarra.es

Abstract

The spectral study of the glottal excitation has traditionally been based on a single time-domain mathematical model of the signal, and the spectral dependence on its time domain parameters. Opposite to this approach, in this work the two most widely used time domain models have been studied jointly, namely the KLGLOTT88 and the LF models. Their spectra are analyzed in terms of their dependence on the general glottal source parameters: Open quotient, asymmetry coefficient and spectral tilt. As a result, it has been proved that even though the mathematical expressions for both models are quite different, they can be made to converge. The main difference found is that in the KLGLOTT88 model the asymmetry coefficient is not independent of the open quotient and the spectral tilt. Once this relationship has been identified and translated to LF model, both models are shown to be equivalent in both time and frequency domains.

1. Introduction

During the last decades, the source-filter model has been adopted by most researchers in the area of speech production. In this context, voice source measurements are needed for high quality speech or singing synthesis, because accurate voicing is currently a key issue for naturalness. It is also known, that the source filter separation process is seldom arbitrary, and that the choice of source parameters will affect the vocal tract estimation and vice versa. However, good frequency matching between the original and the resynthesized signal would ensure a proper estimation of the glottal source and vocal tract parameters.
In most literature on the topic, the voice source estimation problem has traditionally been posed in the time domain, and thus almost all mathematical glottal source models are described by temporal parameters [1, 2, 3]. This is because the meaningful physical glottal source parameters are normally defined in the time domain, this is, Open quotient, asymmetry coefficient, fundamental period, and spectral tilt. This last parameter is the only one whose spectral influence is the most evident [4], being the effect of the others not so.

To cope with this limitation, there have been some research efforts in the direction of analyzing the effect in the spectrum of the time domain parameters of a given mathematical glottal source model. Obviously these time domain parameters are dependent on the mathematical expression of the model, and, depending on the model they do not exactly correspond to the open quotient, asymmetry coefficient or the spectral tilt. For example, in [5, 6] the analytic expression of the spectrum of the KLGLOTT88 and the LF model are provided, and both are studied in terms of its own time domain parameters. As a result of this analysis, some relationships between spectral features (H1-H2) and time domain parameters have been proposed. In [7] the LF model is considered and the spectral influence of the timing parameters, R_g, R_k and R_a is studied, and it is proved that R_a is the most relevant parameter in terms of the spectrum of the LF model.
All of these works have tried to relate somehow the spectral characteristics of a glottal source model with its time domain parameters. Besides, some spectral parameters have been proposed, independently of time domain parameters to parameterize different phonation types. In this way, in [8, 9], a new parabolic spectral parameter has been proposed in order to describe the glottal source. However, it is not well known the relation of this parameter with time domain parameters.
In this context, the two glottal source models KLGLOTT88 and LF, have been jointly studied in this paper, analyzing their spectral dependence on the: open quotient, asymmetry coefficient and spectral tilt. Then, the analytic expressions for the spectra of the two models proposed in [5], have been considered. In a next step it will be shown that the spectrum of both models is quite similar, in spite of the differences between their time-domain description. Then, the asymmetry coefficient of the KLGLOTT88 has been obtained in terms of the open quotient and the spectral tilt. Finally, it will be shown that when this dependence is included in the LF model, both become equivalent in both time and frequency domains.

2. KLGLOTT88 and LF models

In this section, the two mentioned mathematical models are reviewed, and it will be shown how the open quotient, asymmetry coefficient and spectral tilt are included:

2.1. KLGLOTT88 model

In [2] the KLGLOTT88 model was proposed. The block diagram describing this model is shown in Fig.1:

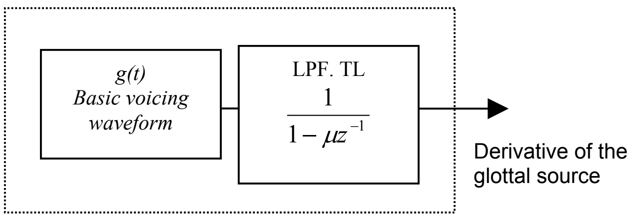


Fig. 1. Block diagram of the KLGLOTT88 model

The first block accounts for the open quotient of the signal, and the amplitude of voicing. Then, the spectral tilt is added by a first order low pass filter. In the literature, the basic waveform can represent either the derivative of the glottal source or the glottal source itself, depending on if the lips radiation effect is included or not. In the following expression, *g(t)* represents the basic voicing waveform of the derivative of the glottal source:

$$g(t)=\begin{cases}2at-3bt^2 & 0\le t\le O_q\cdot T_o\\ 0 & O_q\cdot T_o\le t\le T_o\end{cases} \qquad (1)$$

where

$$a=\frac{27\cdot AV}{4T_oO_q^{\ 2}} \qquad b=\frac{27\cdot AV}{4T_o^{\ 2}O_q^{\ 3}}$$

and being the spectral tilt characterized by the cut-off frequency of the low pass filter, f_c. From this point on, this frequency will be normalized to the fundamental frequency of the signal, and thus it will be expressed as $f_t = f_c/F_0$.

It is evident that in this model the asymmetry coefficient is determined by the mathematical formula of the glottal source, and in this case, it can not be freely modified.

In [5], an analytical study of the spectrum of the KLGLOTT88 model was developed, and there, a mathematical expression was proposed for the continuous spectrum of *g(t)*:

$$G(f)=\frac{27AV}{2O_q(2\pi f)^3}\left[\frac{je^{-j2\pi fO_qT_o}}{2}+\frac{1+2e^{-j2\pi fO_qT_o}}{2\pi fO_qT_o}+ +3j\frac{1-e^{-j2\pi fO_qT_o}}{\left(2\pi fO_qT_o\right)^2}\right] \qquad (2)$$

It is possible to see, that the open quotient controls the spectral shape of (2), and the amplitude of voicing becomes a scale factor.

In Fig. 2.a and Fig. 2.b, the spectral effect of O_q and f_t is depicted, and also, the continuous spectrum of the derivative of the glottal source is represented for several values of O_q and f_t:

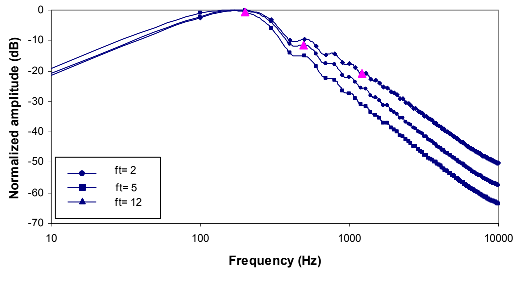


Fig. 2.a Spectral effect of f_t. O_q= 0,4. F_o= 100Hz

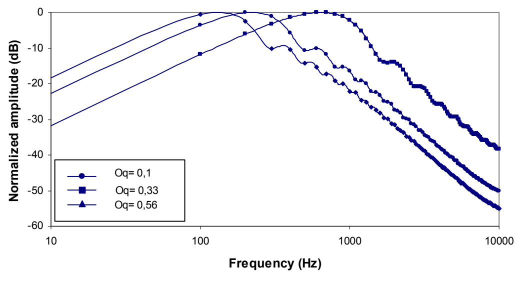


Fig. 2.b Spectral effect of O_q. f_t= 10. F_o= 100Hz

The continuous spectrum of the derivative of the glottal source is characterized by a low frequency parabolic shape, which is related to the spectral parameter presented in [8, 9]; a damped ripple and a spectral roll-off. Looking at Fig. 2.a, it is evident, because of the two block structure of the model, that f_c splits the spectrum of the derivative of the glottal flow in two regions with different slope. Moreover, looking at Fig. 2.b,a relationship between the open quotient and the width of the parabolic region is apparent. The shape of this region will be affected by the spectral tilt depending on the value of f_c, given an open quotient.

2.2. LF model

In [1] the LF model was introduced. The known expression for the derivative of the glottal flow follows:

$$u'(t)=\begin{cases}E_oe^{\alpha t}\sin\omega_g t & 0\le t\le t_e\\ \dfrac{-E_e}{t_a\varepsilon}\left(e^{-\varepsilon(t-t_e)}-e^{-\varepsilon(t_c-t_e)}\right) & t_e\le t\le t_c\end{cases} \qquad (3)$$

The eight direct synthesis parameters are determined by three independent timing parameters, R_g, R_k, R_a, which are related with the relevant time instants, i.e.: T_e, which represents the time instant when the derivative of the glottal source has a negative peak, T_p,, which represents the instant when the glottal flow is maximum, and T_a, which is related with the time constant of the closing phase [1]:

$$R_g=\frac{T_o}{2\cdot T_p} \qquad R_k=\frac{T_e-T_p}{T_p} \qquad R_a=\frac{T_a}{T_o} \qquad (4)$$

This three parameters can be further related to the open quotient, asymmetry coefficient and the normalized cut-off frequency of the spectral tilt as:

$$O_q=\frac{1+R_k}{2\cdot R_g} \qquad \alpha=\frac{1}{R_k+1} \qquad f_t=\frac{1}{2\pi R_a} \qquad (5)$$

Given a set of values for the timing parameters (R_g, R_k, R_a), or a combination of O_q, α, f_t, a system of non linear equations must be solved to generate the direct synthesis parameters of the model according to expression (3), what makes this model cumbersome to be used [1, 3, 5].

As in the case of the KLGLOTT88 model, a mathematical expression for the continuous spectrum of this model was proposed in [5]:

$$U(f)=E_0\frac{1}{(\alpha-j2\pi f)^2+\omega_g^2}\Big[\omega_g+ +e^{(\alpha-j2\pi f)T_e}\left((\alpha-j2\pi f)\sin\omega_gT_e-\omega_g\cos\omega_gT_e\right)\Big]+ +E_e\frac{e^{-j2\pi fT_e}}{\varepsilon T_aj2\pi f(\varepsilon+j2\pi f)}\left[\varepsilon(1-\varepsilon T_a)\left(1-e^{-j2\pi f(T_o-T_e)}\right)-\varepsilon T_aj2\pi f\right] \qquad (6)$$

By comparing (1) with (3) or (2) with (6) it appears that both models are very different. However, it will be shown that this firs appreciation is not completely true.In Fig. 3.a, Fig. 3.b and Fig. 3.c, the spectral effect of the time domain parameters is represented. However, in this model there are three independent parameters: O_q, α, and f_t.

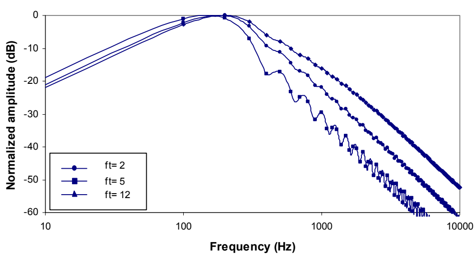


Fig.3.a. Spectral effect of f_t. O_q= 0,4, α= 0,75. Fo= 100Hz

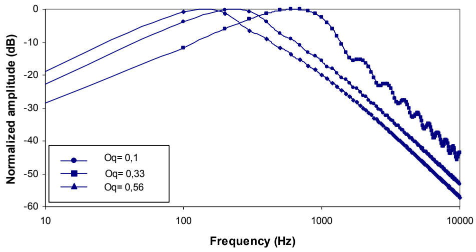


Fig. 3.b. Spectral effect of O_q. f_t= 10, α= 0,75. Fo= 100Hz

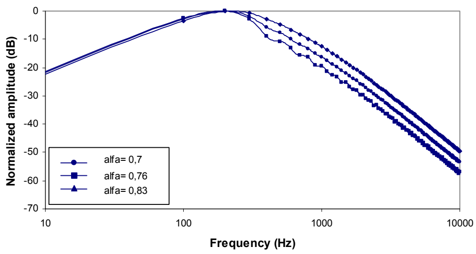


Fig. 3.c. Spectral effect of α. f_t= 10, O_q= 0,4. Fo= 100Hz

By comparing the KLGLOTT88 and the LF spectra it can be seen that they are very similar. Actually, the spectral influence of f_c is very similar in both cases, because it again represents the frequency where the spectrum changes its slope. Regarding the O_q spectral influence, it is possible to observe that as O_q decreases the spectrum widens. However, considering the parabolic low frequency area, and the spectral ripple, it is evident that these features also change with f_c and O_q, in a different way to the KLGLOTT88 model. This is because in the LF model a set of non linear equations has to be solved in order to obtain the direct synthesis parameters from the timing parameters, and therefore, it is reasonable to think that the spectral effect of the three time domain parameters must be related. Besides, there is another degree of freedom in this model, the asymmetry coefficient. In Fig. 3.c the spectral influence of this parameter is represented. For this case, the two-slope shape is maintained; however the low frequency parabolic region and the spectral ripple depend on α.
Thus, although both mathematical models are analytically very different, their spectra are somehow similar, being the main difference caused by the implicit dependence of one parameter on other two in the case of KGLOTT88. This will be shown in the next section.

3. The asymmetry coefficient in the KLGLOTT88 model

In this section, the asymmetry coefficient of the KGLOTT88 model will be characterized. For this model, the time instant when the glottal flow is maximum, T_p, has been calculated, for several values of the O_q and the f_t. This is shown for instance in Fig. 4:

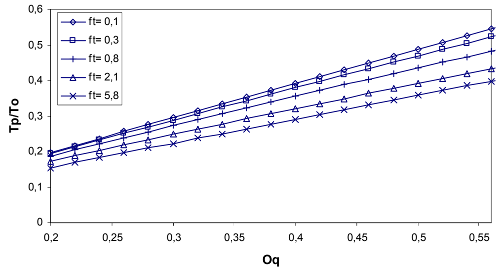
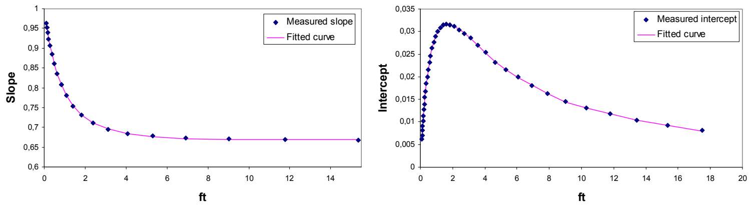


Fig. 4. Maximum glottal flow instant in the KLGLOTT88 model

Looking at Fig. 4, a linear relationship between T_p/T_o and O_q is apparent, a thus it can be modeled as:

$$\frac{T_p}{T_o} = a(f_t) O_q + b(f_t) \qquad (7)$$

being the dependence of slope, $a(f_t)$, and intercept point, $b(f_t)$, as shown in Fig. 5:



(5.a) (5.b)
Fig. 5. Linear parameters of T_p/T_o - O_q. (5.a) Slope (5.b) intercept

These two dependencies can be approximated by the following exponential functions:

$$a(f_t) = a_o + a_1 e^{-\frac{f_t}{A_1}} + a_2 e^{-\frac{f_t}{A_2}} \qquad (8)$$

where a_o= 0,67, a_1= 0,17, a_2= 0,16, A_1= 0,58, A_2= 1,73, and

$$b(f_t) = b_o \cdot \left(1 - e^{-\frac{f_t}{B_1}}\right) \cdot \left(e^{-\frac{f_t}{B_2}} + e^{-\frac{f_t}{B_3}}\right) \qquad (9)$$

where b_o= 0,0231, B_1=0,72, B_2=3,46, B_3=16,23.
Under these conditions the asymmetry coefficient of the KLGLOTT88 model can be expressed as:

$$\alpha_{KLGLOTT88} = \frac{T_p}{O_q T_o} = a(f_t) + \frac{b(f_t)}{O_q} \qquad (10)$$

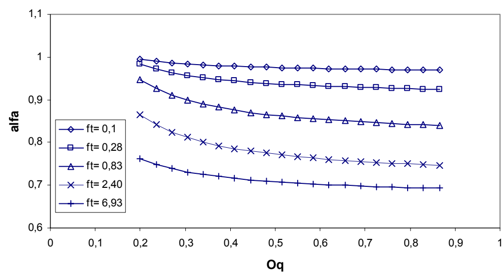


Fig. 6. Asymmetry coefficient in the KLGLOTT88 model

Fig. 6 shows that the asymmetry coefficient is not constant in the KLGLOTT88 model, but it depends on O_q and f_t.

4. LF model as KLGLOTT88 model

In this section, we will compare the spectra of the two models under the same conditions, this is, the LF waveform will be represented using the asymmetry coefficient resulting from expression (10):

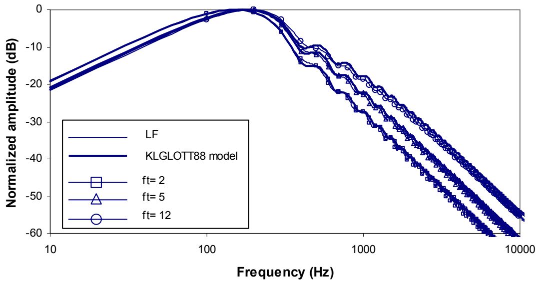


Fig. 7.a. LF and KLGLOTT88 spectra for different f_t. $O_q= 0,4$ $Fo=100Hz$

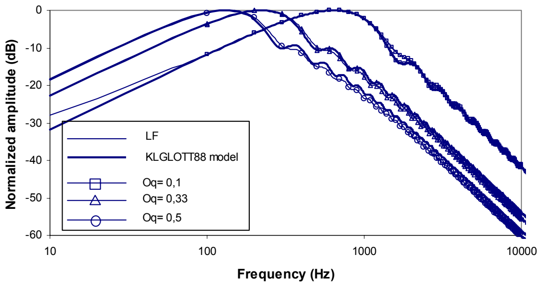


Fig. 7.b. LF and KLGLOTT88 spectra for different O_q. $f_t= 10$ $Fo=100Hz$

Figures 7.a and 7.b show, for both models, how the spectra change versus O_q or f_t variations. As we have seen, asymmetry coefficient is then determined in the KGLOTT model, whereas in the LF it has been forced to be the same for comparison purposes. Apart from minor differences, the pairs of spectra result indistinguishable. Under these conditions, it is possible to say that the KLGLOTT88 model is a particular case of the LF model, where asymmetry coefficient is determined by O_q and f_t.

Considering that the LF model is the most powerful, since it has three degrees of freedom, it has to be said that, in general, the shape of the low frequency parabolic area, and the spectral ripple are determined by the three time domain parameters: open quotient, asymmetry coefficient and the spectral tilt. This statement is somehow represented by Fig. 8:

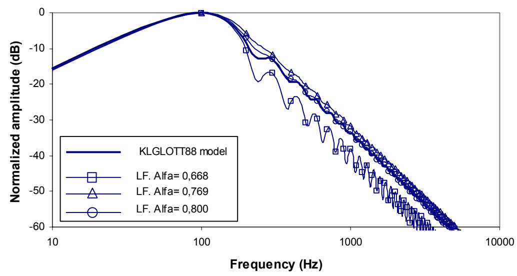


Fig. 8 Asymmetry coefficient in the LF model. $O_q= 0,5, f_t= 5.$ $Fo =100 Hz.$ $\alpha_{KLGLOTT88}= 0,769$

When α is lower than $\alpha_{KLGLOTT88}$ the parabolic region and the spectral ripple are bigger. Otherwise, if α is higher than $\alpha_{KLGLOTT88}$, the spectral ripple decreases.

5. Conclusions

By comparing two different time domain models for the derivative of the glottal source, it is possible to conclude that the continuous spectrum of the derivative of the glottal source is characterized by three main features: a low frequency parabolic area, a spectral ripple and a frequency, which splits the spectrum into two different slope intervals. Taking into account these three characteristics, the KLGLOTT88 model behaves in a simple way, because the spectral ripple remains invariant regardless of its time parameters. The reason behind this spectral behavior is that this model has only two independent parameters: the open quotient and the spectral tilt, being the asymmetry coefficient dependent on the former. On the other side, the LF model allows an independent control of the three main glottal source parameters. In this way, the three spectral features can be modified independently, but when the asymmetry coefficient follows the dependence of the KLGLOTT88 model, both models are shown to be equivalent in time and frequency domains.

From this spectral study, it can be also deduced that the parabolic spectral parameter proposed in [8, 9] is a global source parameter because it depends on the open quotient, asymmetry coefficient and the spectral tilt. Also, a deeper study of this parameter could provide a closed form expression of this dependence, and identify which of the three parameters is the most correlated with this spectral feature.

6. Acknowledgements

The Gobierno de Navarra and the Universidad Pública de Navarra are gratefully acknowledged for financial support.

7. References

[1] G. Fant, J. Liljencrants, Q. Lin, "A four-parameter model of glottal flow", *STL-QPSR*, Vol. 85, nº 2, pp. 1-13, 1985

[2] D. H. Klatt, L. C. Klatt, "Analysis, synthesis and perception of voice quality variations among female and male talkers", *Journal of the Acoustical Society of America*, Vol. 87, nº 2, pp. 820-857, February 1990

[3] R. Veldhuis, "A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation", *Journal of the Acoustical Society of America*, Vol. 103, nº 1, pp. 566-571, January 1998

[4] G. Fant, Q. Lin, "Frequency domain interpretation and derivation of glottal flow parameters", *STL-QPSR*, Vol. 88, nº 2-3, pp. 1-21, 1988

[5] B. Doval, Ch. d'Alessandro, "Spectral correlates of glottal waveform models: An analytic study", *Proceedings of the IEEE ICASSP*, pp. 446-452, April 1997

[6] B. Doval, Ch. d'Alessandro, B. Diard, "Spectral methods for voice source parameters estimation", *Proceedings of the EUROSPEECH*, Vol. 1, pag. 533-536, Rhodes, September 1997

[7] R. N. J. Veldhuis, "The spectral relevance of glottal-pulse parameters", *Proceedings of the IEEE ICASSP*, 1998

[8] P. Alku, E. Vilkman, "A frequency domain method for parametrization of the voice source", *Proceedings of the ICSLP*, 1996

[9] P. Alku, H. Strik, E. Vilkman, "Parabolic spectral parameter - a new method for quantification of the glottal flow", *Speech Communication*, nº 22, pp. 67-79, 1997