GLOTTAL FLOW MODELS : WAVEFORMS, SPECTRA AND PHYSICAL MEASUREMENTS

PACS: 43.70.Bk

Henrich Nathalie; d'Alessandro Christophe; Doval Boris
LIMSI - CNRS
BP 133, Université Paris XI
F-91403 Orsay
France
Tel: (33) 1 69 85 81 13
Fax: (33) 1 69 85 80 88
E-mail: henrich@lam.jussieu.fr; cda@limsi.fr; doval@limsi.fr

ABSTRACT

Most of the main time-domain glottal flow models proposed in the literature can ultimately be described in a unified manner by a set of five parameters: the fundamental frequency f_0, the amplitude of voicing A_v, the open quotient O_q, the asymmetry coefficient α_m and the return phase quotient Q_a. Time and frequency-domain analytical formulations are given in the case of an abrupt closure. The two main spectral components (glottal formant and spectral slope) are detailed and related to the time-domain parameters. The use of signal models for voice source analysis is discussed with regards to the open quotient, on the basis of physical (electroglottographic and acoustic) measurements.

INTRODUCTION

Modelling voice production is a challenging issue in acoustics, both on the theoretical side and on the application side. As pertains to the theory, voice production seems particularly difficult to explore, because the voice apparatus is very complex (the tongue alone comprises about 17 muscles), because the dynamics of speech and singing is very intricate, and last but not least because direct experimental measurements are difficult to obtain as invasive acoustic and physiological measurements are of course excluded on live healthy subjects.
Almost all the studies in voice production assume a source/filter model (Fant[6], 1960). At least for vowels, the voice source signal is produced at the glottis, and then filtered by (passive) vocal tract and sound radiation components. Several types of voice source models can be envisaged for studying vocal production, and have been more or less developed so far:

- Fluid mechanics models: the most general type of model is based on fluid mechanics. 3-D Navier-Stockes equations have to be solved for computing the air flow in the glottis and vocal tract. Sound is only the compressible part of the model. This type of model is currently studied in several research teams, but because of the computing power needed, the results obtained so far are still very limited. This type of model may help to pass beyond the limitations of the source-filter hypothesis (Ramsay[21], 2001).
- Mechanical and electroacoustic models: the prototype of this model is the two-mass model (Ishizaka and Flanagan[16], 1972) recently revisited by Pelorson et al.[20] (1994). A mechanical model of the voice production system (in terms of springs and mass) is coupled with a delay line representation of the vocal tract (using electroacoustic analogies).

- Signal models: assuming a source-filter decomposition and plane-wave propagation, the acoustics of voice production is reduced to 1-D signal processing. The advantage of the signal approach is that the parameters obtained can be linked both to production and to perception. This is the approach used in a large majority of voice processing applications, and the approach developed here.

In this paper we present a review of our recent work on voice source modelling. Voice quality is one of our main concerns, both for singing voice and for speech. The work on voice source models is conducted for dealing with for example: the acoustic study of registers in speech and singing, characterisation of pressed vs. relaxed voice quality, and modelling of the vocal effort. According to our current knowledge, only signal models are able to deal with these voice-quality related matters. In the first section, time-domain voice source (glottal flow) models are reviewed, and a unified set of parameters is proposed. The spectrum of these models is described in the second section, and a set of spectral parameters is also proposed. In the third section, signal models are compared to physiological and acoustic measurements with regards to the open quotient. The last section concludes with a discussion on the current uses of signal models for voice source processing, including ideas on future work and the model limitations.

UNIFICATION AND ANALYTICAL EQUATIONS.

A unified set of parameters

Several time-domain glottal flow models have been proposed in the literature (Rosenberg[22], 1971; Fant[7], 1979; Hedelin[12], 1984; Fant et al.[8], 1985; Ljungqvist[18], 1986; Klatt & Klatt[17], 1990; Milenkovic[19], 1993; Childers and Hu[2], 1994; Fant[9], 1995; Veldhuis[25], 1998). They all share the following common features : they are bell-shaped, positive or null, quasi-periodic, continuous, and differentiable (except at glottal closure in some situations). Nevertheless, they use neither the same parameters nor the same number of parameters.

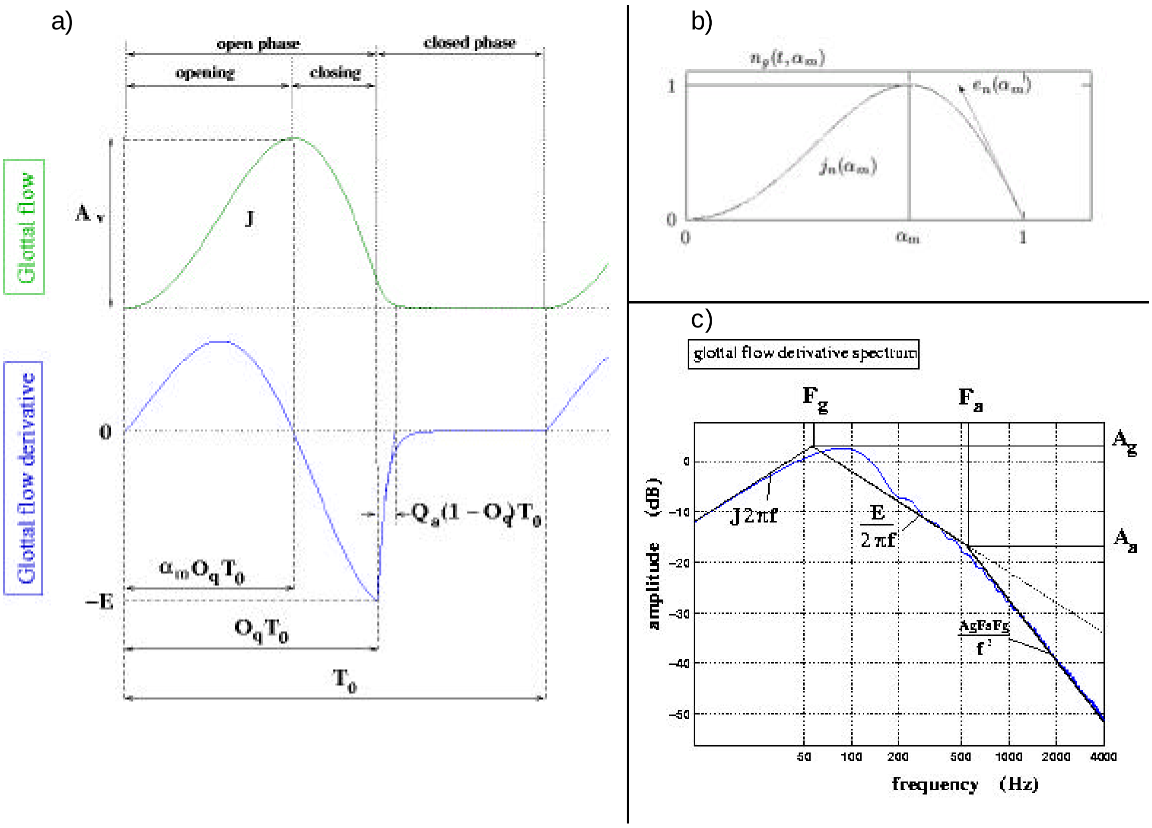

Figure 1: Panel a) : representation of the glottal flow parameters, on one period of the glottal flow and its derivative. A_v, amplitude of voicing; J, integral of the pulse; E, negative peak amplitude of the differentiated glottal flow; T_0, fundamental period; O_q, open quotient; α_m, asymmetry coefficient; Q_a, return phase quotient. Panel b) : normalised glottal flow model (e_n, normalised maximum excitation; j_n, integral of the pulse). Panel c) : glottal flow derivative spectrum.

Thus, a work of unification of these models has been conducted by Doval & d'Alessandro[4],[5] (1997, 1999). It was shown that all glottal flow models can ultimately be described by a set of five time-domain parameters, which are shown in Figure 1 (panel a):

- the fundamental period (T_0) or its equivalent, the fundamental frequency (f_0);
- the amplitude of voicing (A_v);
- the open quotient (O_q), defined as the ratio between the glottal open time and the fundamental period;
- the asymmetry coefficient (α_m), defined as the ratio between the flow rise time (from baseline to peak flow) and the open time [1];
- the return phase quotient (Q_a), defined as the ratio between the effective return phase duration and the closed phase duration $(1-O_q)T_0$. In the case of an abrupt closure of the vocal folds, $Q_a = 0$;

<u>Time-domain equations</u>

Conversion of the original set to this unified set of parameters has been demonstrated by Doval & d'Alessandro for four models: Rosenberg C [22] (1971), LF model [8] (1985), KLGLOTT88 model [17] (1990) and R++ model [25] (1998). In the case of an abrupt closure, they have found that the time-domain equations describing the glottal pulse train $u_g(t)$ and its derivative $u'_g(t)$ for a given model can be rewritten as follows:

$$u_g(t) = A_v\, n_g\left(\frac{t}{O_q T_0}, \alpha_m\right) * Ш_{T_0}(t)$$

$$u'_g(t) = \frac{A_v}{O_q T_0}\, n'_g\left(\frac{t}{O_q T_0}, \alpha_m\right) * Ш_{T_0}(t)$$

where $Ш_{T_0}(t)$ is a Dirac comb filter with fundamental period T_0. The function $n_g(\xi,\alpha_m)$ is a normalised model-based function, which depends only on the pulse asymmetry, as shown in Figure 1 (panel b). The amplitude of voicing (A_v) can easily be replaced by another parameter, the negative peak amplitude of the differentiated glottal flow (E), in applying the following relation :

$$E = \frac{A_v}{O_q T_0}\, e_n(\alpha_m)$$

<u>Frequency-domain equations</u>

The spectral equations can be derived in applying a Fourier transform to $u_g(t)$ and $u'_g(t)$:

$$U_g(f) = A_v O_q T_0\ N_g(f O_q T_0, \alpha_m)\ (f_0 Ш_{f_0}(f))$$

$$U'_g(f) = A_v\ N'_g(f O_q T_0, \alpha_m)\ (f_0 Ш_{f_0}(f))$$

where $Ш_{f_0}(f)$ is a Dirac comb filter with fundamental frequency f_0. $N_g(\xi,\alpha_m)$ and $N'_g(\xi,\alpha_m)$ are the Fourier transforms of $n_g(\xi,\alpha_m)$ and $n'_g(\xi,\alpha_m)$.

With the help of these analytical equations, we can explore the glottal flow spectral correlates.

SPECTRAL DESCRIPTION

<u>Glottal formant</u>

A second-order low-pass filter can approximate the glottal flow spectrum. This spectral reinforcement is called *glottal formant* [2] (Fant[7], 1979). Two asymptotic lines are defined: the first with a *0 dB/oct* slope when frequency tends toward 0 [$(D_1) = J$], and the second with a *–12 dB/oct* when frequency tends to infinity [$(D_2) = E/(2\pi f)^2$]. In case of the glottal flow derivative, the asymptotic lines have slopes of *+6 dB/oct* [$(D_1') = J\, 2\pi f$] and *–6 dB/oct* [$(D_2') = E/(2\pi f)$], as

---

[1] This parameter is related to the so-called "speed quotient" (S_q), defined as the ratio between the flow rise time and the flow fall time (from peak flow back to baseline): $\alpha_m = S_q/(1+S_q)$. It was introduced to simplify the equations of glottal flow models.

[2] As the term "glottal formant" may be confusing, we should precise it. It is used to design a spectral maximum in the glottal source spectrum. The glottal formant is not related to any resonance phenomenon.

shown on Figure 1 (panel c). The cut-off frequency F_g and amplitude A_g can be computed from the coordinates of the crossing point between the two asymptotic lines. They are related to the time-domain parameters by the following equations:

$$F_g = \frac{1}{2\pi O_q T_0}\sqrt{\frac{e_n(\alpha_m)}{j_n(\alpha_m)}} = \frac{1}{2\pi}\sqrt{\frac{E}{J}}$$

$$A_g = A_v\sqrt{e_n(\alpha_m)\, j_n(\alpha_m)} = \sqrt{E\ J}$$

where $e_n(\alpha_m)$ and $j_n(\alpha_m)$ are two model-based functions defined on Figure 1 (panel b), which depend only on the asymmetry of the pulse. The glottal formant frequency is thus dependent on the 3 parameters f_0, O_q and α_m. Its behaviour relative to the fundamental frequency is shown in Figure 2 in the case of two models: a model where the asymmetry is a constant (KLGLOTT88) and another where the asymmetry coefficient can be varied (LF model). The higher the asymmetry of the pulse or the lower the open quotient, the higher the frequency of the glottal formant. Around a typical value of asymmetry coefficient ($\alpha_m \approx 2/3$) and for normal values of open quotient ($O_q \geq 0.5$), the glottal formant is located slightly below or close to the first harmonic ($H_1 = f_0$). It can reach for instance the fourth harmonic when $O_q = 0.4$ and $\alpha_m = 0.9$. Thus, the glottal flow can contribute to the spectral enhancement of precise harmonics in the low-frequency part of a voiced signal spectrum. This contribution is mainly due to the open quotient and the asymmetry coefficient.

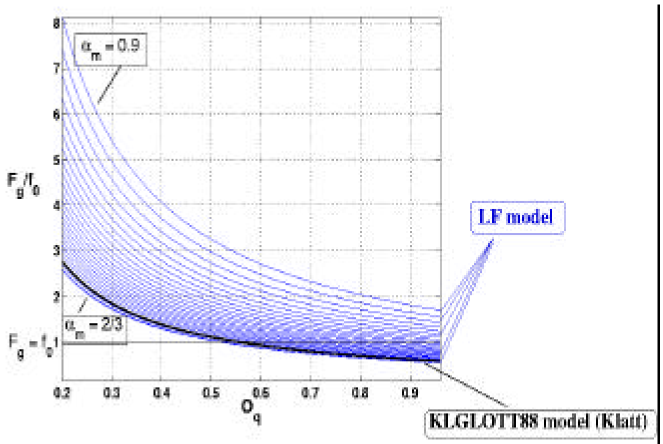

Figure 2: Spectral position of the glottal formant cut-off frequency, relative to the fundamental frequency, as a function of open quotient and asymmetry coefficient.

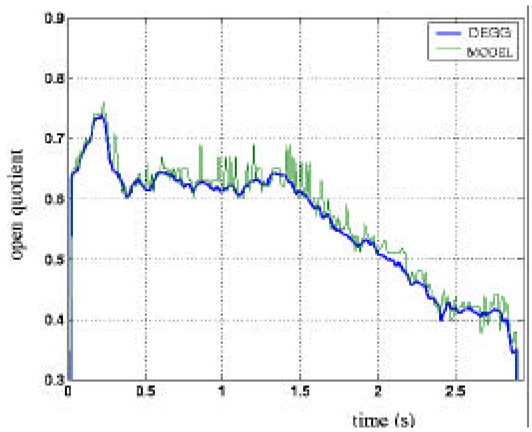

Figure 3: comparisons of open quotient measurements on a relaxed-to-pressed male phonation. DEGG: measures on the derivative of electroglottographic signals; MODEL: model-based estimation on the acoustic signal, using a pitch-synchronous covariant linear prediction method (Ljungqvist[18]).

<u>Spectral slope</u>

Up to now, we have assumed an abrupt closure. In this case, the spectral slope depends only on the variation of the parameter E, which can be induced by variations of other parameters, such as the asymmetry coefficient. When E is kept constant, the spectral slope keeps a fixed spectral position. A smooth closure of the vocal folds implies that the parameter Q_a is not null anymore. This effect can directly be introduced in time-domain models, by adding a decreasing exponential to the glottal flow derivative during the closed phase (Veldhuis[25], 1998). In the spectral domain, a first-order low-pass filter with a high cut-off frequency can model it. It corresponds to the addition of a third asymptotic line with a *–12 dB/oct* slope [$(D'_3) = A_g F_a F_g / f^2$], as shown in Figure 1 (panel c). According to Fant [8],[9], its cut-off frequency F_a is related to the time-domain parameters by the following equation:

$$F_a \approx \frac{1}{2\pi Q_a (1 - O_q T_0)}$$

Its amplitude A_a is then given by:

$$A_a = \frac{E}{2\pi F_a} = \frac{F_g}{F_a} A_g$$

The five time-domain parameters can thus be easily replaced by these frequency-domain parameters, which allows a spectral description and analysis of the glottal flow.

PHYSICAL MEASUREMENTS

Do the parameters of these signal models have a physical meaning? Can a glottal flow model reproduce all the variability of a real signal? In order to answer to these questions physiological and acoustical measurements have been compared. We will focus here on open quotient, as it can easily be measured on an electroglottographic signal or estimated by inverse-filtering and spectral analysis.

Electroglottographic measurements

Physiological measurements of open quotient can be made on the derivative of electroglottographic (EGG) signals (Henrich[13]). The EGG signal is proportional to the vocal folds' contact area. Its derivative shows strong peaks at the glottal closure instants and weak peaks with opposite sign at the glottal opening instants (Childers et al.[3]). Thus, the open quotient can be calculated as the ratio between the duration from a glottal opening instant to the following glottal closure instant (open time) and the duration between two glottal closure instants (fundamental period). The comparisons between these measures and acoustical model-based estimation of open quotient were conducted on sustained vowels and relaxed-to-pressed male phonation. As illustrated in Figure 3, a good agreement is found between both measurements, which confirms the results obtained by Childers et al.[1] (1990).

Spectral estimation

Several experimental studies have shown a strong correlation between the open quotient and the amplitude difference $(H_1\text{-}H_2)$ between the two first harmonics of the glottal flow derivative spectrum (Holmberg et al[15], Fant[9],[10], Stevens & Hanson[23], Hanson[11], Sundberg et al.[24]). An empirical formulation of this correlation has been proposed by Fant[9],[10] on the basis of speech measures:

$$(H_1 - H_2) = -6 + 0{,}27 * e^{5{,}5Oq}$$

The theoretical values of *(H1-H2)* can be calculated as a function of the glottal flow parameters, in using the spectral formulations of glottal flow models. It has been shown that this amplitude difference is not only dependent on open quotient, but also on asymmetry coefficient and return phase quotient (Henrich et al.[13],[14]). As it is only dependent on open quotient with regards to Fant's empirical formulation, a resulting dependency between glottal flow parameters can be computed. The results show that a low value of open quotient $(0.25 \le Oq \le 0.6)$ is correlated to a high value of asymmetry coefficient ($\alpha_m \ge 0.7$) and a low value of return phase quotient ($Q_a \le 0.05$). In the same way, a high value of open quotient ($Oq \ge 0.6$) is correlated to a low value of asymmetry coefficient ($0.55 \le \alpha_m \le 0.7$) and higher values of return phase quotient are then possible. These results are in agreement with the acoustic correlates of pressed and relaxed phonation. A pressed phonation is characterised by a long closed phase, an increased pulse skewing and an abrupt return phase. On the contrary, a relaxed phonation is characterised by a rather short closed phase, a symmetrical pulse and a smooth closure.

In the case of speech, a good agreement is thus found between theoretical prediction and experimental measurements. This is no longer true when analysing singing voice production. In this case, it was found that the spectral measurements could not be explained by any glottal flow parameters' configuration (Henrich et al.[13],[14]). When there is a strong interaction between source and filter (which can presumably be a main characteristic of singing), inverse filtering may no longer be valid and the usefulness of glottal flow models is then questionable.

CONCLUSION

As compared to present physical models, the main potential of signal models is their usefulness in speech voice-quality related studies. In this context, the unified framework briefly presented here is a tool for understanding the temporal and spectral correlates of the glottal flow. As voice quality is better described by spectral parameters, the development of an analytical spectral formalism is of greater value for perceptual studies of the voice source. Evidences of the physical meaning of glottal flow models' parameters were given with regards to the open quotient, in comparing acoustical and physiological measurements with model-based estimation and prediction. Nevertheless, the usefulness of present glottal flow models is limited by the underlying source-filter independency hypothesis. New approaches should now be developed that would take into account this possible source-filter dependency and would help gain better insight into the properties of voice production.

BIBLIOGRAPHICAL REFERENCES

[1] Childers D.G., Hicks D.M., Moore G.P., Eskenazi L. and Lalwani A.L. (1990) "Electroglottography and vocal fold physiology". J. Speech Hear. Res. 33, pp. 245-254.
[2] Childers D.G. and Hu T. H. (1994) "Speech synthesis by glottal excited linear prediction". J. Acoust. Soc. Am. 96(4), pp. 2026-36.
[3] Childers D.G., Naik J.M., Larar J.N., Krishnamurthy A.K. and Moore G.P. (1983) Electrog-lottography, speech, and ultra-high speed cinematography, Vocal Fold Physiology and Bio-physics of Voice, edited by I.R. Titze and R. Scherer (Denver Center for the Performing Arts, Denver), pp. 202-220, 1983.
[4] Doval B. and d'Alessandro C. (1997) "Spectral correlates of glottal waveform models : an analytic study". In IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Munich, Germany, Apr. 1997, pp. 446-52.
[5] Doval B. and d'Alessandro C. (1999) "The spectrum of glottal flow models." Notes et documents LIMSI 99-07.
[6] Fant G. (1960) "Acoustic theory of speech production". Mouton, La Hague.
[7] Fant G. (1979) "Glottal source and excitation analysis". STL-QPSR 1, pp. 85-107.
[8] Fant G., Liljencrants J. and Lin Q. (1985) "A four-parameter model of glottal flow". STL-QPSR 4, pp. 1-13.
[9] Fant G. (1995) "The LF-model revisited. Transformations and frequency domain analysis". STL-QPSR 2-3, pp. 119-56.
[10] Fant G. (1997) "The voice source in connected speech". Speech Communication 22, pp. 125-39.
[11] Hanson H.M. (1997) " Glottal characteristics of female speakers : Acoustic correlates". J. Acoust. Soc. Am. 101, pp. 466-81.
[12] Hedelin P. (1984) "A glottal LPC-vocoder". In IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, pp. 161-64.
[13] Henrich N. (2001) "Etude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électroglottographiques, perception". PhD thesis, Université Paris 6.
[14] Henrich N., d'Alessandro C. and Doval B. (2001) "Spectral correlates of voice open quotient and glottal flow asymmetry : theory, limits and experimental data". In Proc. EUROSPEECH 2001, Aalborg, Denmark, Sept. 01.
[15] Holmberg E.B., Hillman R.E., Perkell J.S., Guiod P.C. and Goldman S.L. (1995) "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice". J. Speech Hear. Res. 38, pp. 1212-23.
[16] Ishizaka K. and Flanagan J.L. (1972) "Synthesis of voiced sounds from a two mass model of the vocal cords". Bell Syst. Tech. Journal 50, 6, pp. 1233-68.
[17] Klatt D. and Klatt L. (1990) "Analysis, synthesis, and perception of voice quality variations among female and male talkers". J. Acoust. Soc. Am. 87(2), pp. 820-57.
[18] Ljungqvist M.G. (1986) "Speech analysis-synthesis based on modeling of voice source and vocal tract characteristics". PhD thesis, University of Tokyo.
[19] Milenkovic P.H. (1993) "Voice source model for continuous control of pitch period". J. Acoust. Soc. Am. 93(2), pp. 1087-96.
[20] Pelorson X., Hirschberg A., van Hassel R.R., Wijnands A.P.J., Auregan Y. (1994) "Theoretical and experimental study of quasi-steady flow separation wthin the glottis during phonation. Application to a modified two-mass model." J. Acoust. Soc. Am. 96, pp. 3416-3431.
[21] Ramsay G. (2001) "A quasi-one-dimensional model of aerodynamic and acoustic flow in the time-varying vocal tract: source and excitation mechanisms". In Proc. EUROSPEECH 2001, Aalborg, Denmark, Sept. 01.
[22]Rosenberg A.E. (1971) "Effect of a glottal pulse shape on the quality of natural vowels". J. Acoust. Soc. Am. 49, pp. 583-90.
[23] Stevens K.N. and Hanson H.M. (1995) "Classification of glottal vibration from acoustic measurements". In Vocal Fold Physiology : Voice quality control, O. Fujimura and M. Hirano, Eds., Singular Publishing Group, San Diego, California, pp. 147-70.
[24] Sundberg J, Andersson M. and Hultqvist C. (1999) "Effects of subglottal pressure variation on professional baritone singers' voice sources". J. Acoust. Soc. Am. 105, pp. 1965-71.
[25] Veldhuis R. (1998) "A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation". J. Acoust. Soc. Am. 103, pp. 566-71.