

# GLOTTAL OPEN QUOTIENT ESTIMATION USING LINEAR PREDICTION

Nathalie Henrich (1)(2), Boris Doval (1), Christophe d'Alessandro (1)

(1) LIMSI-CNRS, BP 133, F91403 Orsay, France.

(2) LAM (UPMC, CNRS, Ministère de la culture), 11 rue de Lourmel, 75015 Paris, France.

E-mail : {henrich,doval,cda} @ limsi.fr

## Abstract

A new method for the estimation of the voice open quotient is presented. Assuming abrupt glottal closures, the glottal flow waveform is considered as the impulse response of an anticausal two-poles filter. It is defined by four parameters :  $T_0$ ,  $A_v$ ,  $O_q$  and  $\alpha_m$ . The last three ones are estimated by a second-order linear prediction of the inverse filtered speech. Results on synthetic and natural speech signals are reported and compared with measurements on the corresponding electroglottographic signals.

## 1 Introduction

Analysis of voice source's acoustic parameters is a challenging issue in the domains of speech communication (e.g. speech analysis and synthesis) or speech and voice pathology research. Vocal fold's vibration is responsible for voice and speech quality. Direct measurement of the glottal activity is still difficult, and thus many methods for voice analysis are based on processing of the acoustic signal.

According to the linear source/filter theory of speech production [1], the source signal is derived from the speech pressure waveform with the help of inverse filtering [2]. The glottal volume velocity waveform is obtained from the speech signal by cancelling the effects of the vocal tract resonances. Several time domain glottal flow models have been

proposed, for representation of the glottal flow waveform with a few parameters. All these models are surprisingly close, even if they do not use the same mathematical functions, the same number of parameters, or the same name for similar parameters. This is because they are sharing common features : they are all bell-shaped, positive or null, quasi-periodic, continuous, and differentiable (except at glottal closure in some situations). The most important time-domain parameters for description of glottal flow models are :

1.  $A_v$  : the maximum amplitude of the glottal flow
2.  $T_0$  : the fundamental period
3.  $O_q$  : the open quotient (ranging between 0 and 1), which defines the glottal closure instant, relative to  $T_0$  (it is at time  $t = O_q T_0$ )
4.  $TL$  : the spectral tilt factor. This parameter is linked to the abruptness of glottal closure, thus to the discontinuity of the glottal flow derivative at closure. As a discontinuity in the derivative of a function can change dramatically its spectrum, the abruptness at glottal closure results in a change of spectral tilt in the voice spectrum.
5.  $\alpha_m$  : the asymmetry coefficient (ranging between 0 and 1). This quotient defines

the instant of maximum of the glottal flow, relative to  $T_0$  and  $O_q$  (this instant is at time  $T_m = \alpha_m O_q T_0$ ). The glottal opening phase is always longer than the glottal closure phase. Thus, most models restrict the range of  $\alpha_m$  to  $[0.5, 1]$ , with typical values around 0.6 or 0.7. Some models do not use this parameter.

These parameters are independent, and can be interpreted as follows. The amplitude of voicing  $A_v$  is a global parameter controlling the general amplitude of the source waveform, and thus of the speech wave. The fundamental period  $T_0$  is also a global parameter, which controls the speech melody. The spectral tilt factor, or glottal closure parameter, controls the spectral richness of the voice. It affects mostly the medium and high frequencies of the speech spectrum (say over 1 or 2 kHz). In time domain, this parameter is very local, because it describes only the shape of the glottal waveform at the glottal closure instant. Spectral tilt plays an important role in variations of vocal effort [3], for instance in stressed vs unstressed syllables, or in loud vs soft speech. The open quotient  $O_q$  is defined as the ratio between the duration of the glottis open phase over the fundamental period. This parameter plays also an important role in variations of vocal effort ([4], [5]), but it affects mostly the low frequencies of the speech spectrum (say below 1 kHz).

The main perceptual correlate of open quotient is often described as voice “pressure”. A “pressed” voice corresponds to a small open quotient and a “relaxed” voice to a large one. It happens also that the open quotient equals 1, i.e. when vocal fold closure is incomplete. The asymmetry parameter  $\alpha_m$  controls the shape of the glottal waveform. Like the open quotient, the spectral effect of this parameter is rather limited to low frequencies. A variation of open quotient often involves a variation of asymmetry. A small open quotient corresponds to a large asymmetry, and conversely, a large open quotient corresponds to a small asymmetry. Thus

some models (see e.g. [4, 6]) make no use of a separate asymmetry parameter, and the open quotient is the only parameter defining the glottal waveform shape. However, it seems that two degrees of freedom exist for the glottal waveform shape, and most models are using two parameters, that can be interpreted as an open quotient and an asymmetry parameter.

Estimation of parameters related to the open quotient using spectral processing has been proposed by several authors. Hanson [7] measured the magnitudes of the first (H1) and the second (H2) harmonics, with a correction according to the first formant. She showed a correlation of this measure with open quotient. Alku et al. [6] presented the parabolic spectral parameter, which represents the rate of decay of the low frequencies for the inverse-filtered pitch-synchronous spectrum. Again this parameter is related to different phonation types.

In this paper, a new method for estimation of open quotient and asymmetry is presented. This method is based on a spectral representation of the glottal flow. The glottal flow component is modeled as the truncated impulse response of an anticausal two-pole filter (see [8]) (Section 2). Then open quotient and asymmetry can be estimated from inverse-filtered speech using a second-order linear predictive analysis (Section 3). Experiments on synthesized and natural voiced signals are presented, and an evaluation with the help of electroglottographic (EGG) signals is reported in Section 4. Section 5 concludes this work.

## 2 A linear glottal flow model

Glottal flow models are usually defined in time-domain, like for instance the models proposed by Rosenberg (1971, [9]), Klatt and Klatt (KLGLOTT88 model, 1990, [4]), Fant, Liljencrants & Lin (LF model, 1985, [10]), Veldhuis (R++ model, 1998, [11]). A unified

view of these models is given in [12]. It is shown that they are built on similar principles and therefore can be represented using a unified set of parameters. Figure 1 displays the waveforms obtained for these 4 models, using the same parameters for all the models and assuming an abrupt glottal closure (i.e. minimum spectral tilt). The resulting waveforms are very close together. The simplification induced by the hypothesis of minimum spectral tilt is useful, because the open quotient is well defined only in case of abrupt closure. Moreover, it is realistic to first consider the situation of abrupt closure, because this situation is the most common in speech and singing.

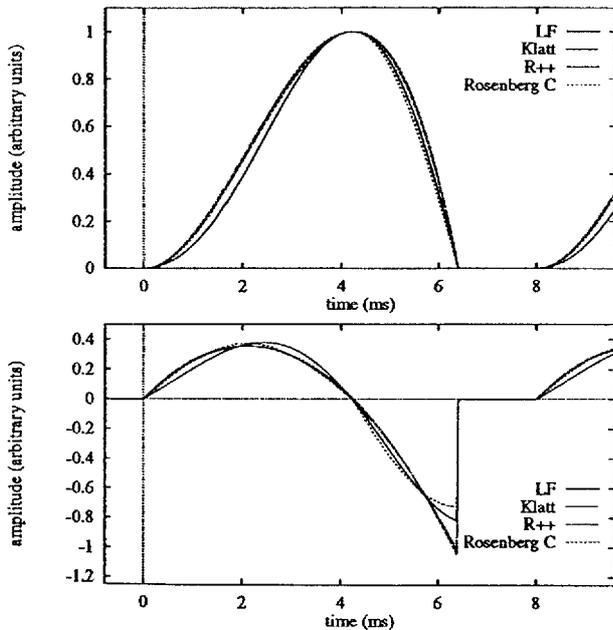


Figure 1: Glottal flow models (top) and their derivatives (bottom) with abrupt closure. (period  $T_0 = 8ms$ , open quotient  $O_q = 0.8$ , asymmetry coefficient  $\alpha_m = 0.66$ )

In [8], analytic formulas for the spectrum of glottal waveforms were computed. It was shown that the glottal flow spectrum was close to a low-pass filter, with order 2 or 3 depending on the spectral tilt component. Assuming minimum spectral tilt, the glottal waveform is actually close to the impulse response of a second-order low-pass filter, provided that time is reversed and that the im-

pulse response is time limited. Thus, if one wants to preserve the glottal pulse shape, it is necessary to design an anticausal filter. If one wants to preserve the finite duration property of the glottal pulse, it is necessary to truncate the impulse response of the filter. Then one can compute an anticausal 2-pole filter, whose impulse response would well approximate the glottal flow. This filter should be described by the set of time-domain parameters mentioned above. The impulse response of a second-order causal filter is :

$$h_c(t) = A \exp(-Bt) \sin(Ct)u(t)$$

where  $u(t)$  is the step function :

$$u(t) = \begin{cases} 1 & \text{if } t > 0 \\ 0 & \text{if } t < 0 \end{cases}$$

Thus, the corresponding anticausal filter is given by :

$$h_a(t) = A \exp(Bt) \sin(-Ct)u(-t)$$

As  $U_g(t)$  opens at time 0 and closes at time  $O_q T_0$ , then  $h_a(t)$  must be shifted by a factor  $\gamma = O_q T_0$  :

$$U_g(t) = A \exp(B(t-\gamma)) \sin(-C(t-\gamma))u(1-\frac{t}{\gamma})$$

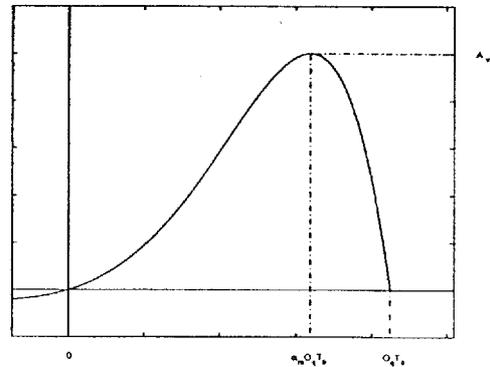


Figure 2: General form of the linear model ( $U_g(t)$ )

The general form of  $U_g(t)$  is presented in Figure 2. The constants A, B and C can be

linked to the model parameters by using the definition of the glottal flow waveform. As  $U_g(0) = 0$  and as  $\alpha_m O_q T_0$  defines the glottal flow maximum instant,  $U'_g(\alpha_m O_q T_0) = 0$  and  $U_g(\alpha_m O_q T_0) = A_v$ , then :

$$\begin{aligned} C &= \frac{\pi}{\gamma} \\ B &= -\frac{\pi}{\gamma \tan(\pi \alpha_m)} \\ A &= \frac{A_v \exp(\frac{\pi(\alpha_m - 1)}{\tan(\pi \alpha_m)})}{\sin(\pi \alpha_m)} \end{aligned}$$

The waveform for one period of the linear glottal flow model is given by :

$$U_g(t) = A_v \frac{\sin(\frac{\pi t}{\gamma})}{\sin(\pi \alpha_m)} \exp(\frac{\pi(\alpha_m - \frac{t}{\gamma})}{\tan(\pi \alpha_m)}) u(1 - \frac{t}{\gamma})$$

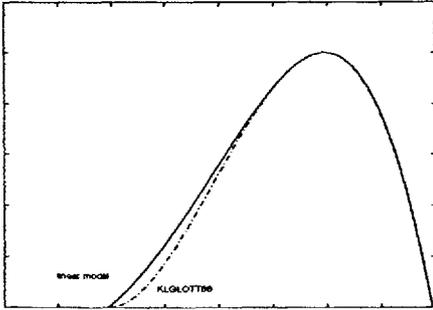


Figure 3: Comparison between the KLGLOTT88 model (dotted lines) and the linear model with  $\alpha_m = 0.7$  (plain lines)

This model is compared to KLGLOTT88 model in Figure 3. The transfer function of the glottal filter, sampled with period  $T_e$ , is given by :

$$\tilde{U}_g(z) = \frac{G_1 z^{-N+1}}{1 + b_1 z + b_2 z^2}$$

where  $(G_1, b_1, b_2)$  are linked to  $(T_0, A_v, O_q, \alpha_m)$  by the following equations :

$$\begin{aligned} b_1 &= -2 \cos(\frac{\pi T_e}{\gamma}) \exp(\frac{\pi T_e}{\tan(\pi \alpha_m) \gamma}) \\ b_2 &= \exp(\frac{2\pi T_e}{\tan(\pi \alpha_m) \gamma}) \\ G_1 &= A_v \frac{\sin(\frac{\pi T_e}{\gamma})}{\sin(\pi \alpha_m)} \exp(\frac{\pi(\alpha_m - 1 + \frac{T_e}{\gamma})}{\tan(\pi \alpha_m)}) \end{aligned}$$

These equations correspond to an infinite impulse response filter. A truncated version of this discrete-time impulse response  $U_g(n)$  is actually used for synthesis. Truncation corresponds to a modulation of the impulse response by a time gate. In the spectral domain, this involves a convolution of the frequency response of the filter by a Sinc function. The main effect of this spectral convolution is to introduce zeros in the spectrum. This seems not very significant for parameter estimation, thus the effect of truncation will not be considered herein.

## 3 Estimation of open quotient

### 3.1 Method

The algorithm for the open quotient estimation is described in Figure 4. According to the source-filter theory [1], the glottal excitation  $\tilde{U}_g(z)$  is filtered by the vocal tract  $\tilde{V}(z)$  and by the lip radiation component  $\tilde{L}(z)$  :  $\tilde{S}(z) = \tilde{U}_g(z) \tilde{V}(z) \tilde{L}(z)$ . Thus the effects of vocal tract and lip radiation must be cancelled from the speech signal for processing the source signal.

An estimation of  $\tilde{V}(z)$  can be obtained by linear prediction of the signal after accentuation : the joint effects of glottal flow and lip radiation are then cancelled. Integration of the speech signal cancels the effect of lip radiation, and thus gives  $\tilde{U}_g(z) \tilde{V}(z)$ . An estimation of the glottal flow is then obtained by filtering  $\tilde{U}_g(z) \tilde{V}(z)$  by the inverse filter of  $\tilde{V}(z)$ .

The theory predicts that the glottal flow can be considered as a second-order system.

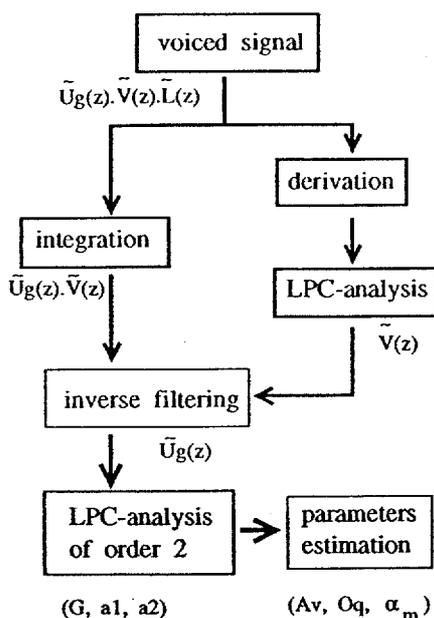


Figure 4: Inverse filtering for glottal parameters' estimation

Therefore, a second-order linear prediction is performed on the estimation of the glottal flow signal, and the coefficients of an optimal 2-pole filter  $F(z) = \frac{G}{1+a_1z^{-1}+a_2z^{-2}}$  are obtained. The autocorrelation method for linear prediction is used, because this method is not sensitive to the direction of time [13]. As the autocorrelation coefficients are symmetric with respect to time, the same linear prediction coefficients are obtained for a causal or an anticausal signal with the same spectral content. Then  $F(z) = \tilde{U}_g(z^{-1})$ , and  $(G, a_1, a_2) = (G_1, b_1, b_2)$ . The glottal parameters and the linear prediction coefficients are related by :

$$O_q = \frac{\pi T_e}{T_0} \frac{1}{\arccos(-\frac{a_1}{2\sqrt{a_2}})}$$

$$\alpha_m = \frac{1}{\pi} [\pi + \arctan(\frac{2\pi T_e}{\gamma \ln a_2})]$$

$$A_v = G \frac{\sin(\pi \alpha_m)}{\sin(\frac{\pi T_e}{\gamma})} \exp(\frac{\pi(1 - \alpha_m - \frac{T_e}{\gamma})}{\tan(\pi \alpha_m)})$$

## 4 Results

In this section, experiments on the estimation of  $O_q$  are reported. This parameter is common to all the models mentioned before. On the contrary,  $\alpha_m$  is fixed in a model such as KLGLOTT88. Moreover,  $O_q$  can be measured experimentally with the help of EGG signals.

Tests were conducted first using synthetic vowel signals ( $F_e = 16$  kHz, order 18 LPC for estimation of the vocal tract).  $O_q$  was chosen to range from 0.4 to 0.9, and  $F_0$  from 80 to 300 Hz. The results for a French vowel /a/, synthesized using KLGLOTT88 model, are shown in Figure 5. Open quotient seems well estimated for a wide range of parameter variation.

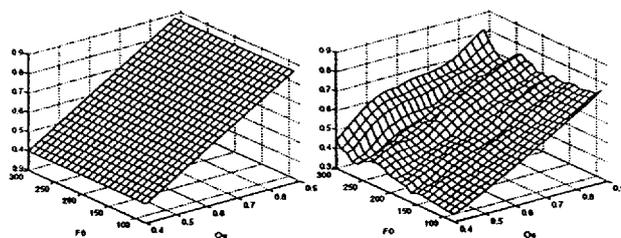


Figure 5: Estimation of  $O_q$ . Synthetic vowel /a/, using KLGLOTT88 source model, for different values of  $O_q$  and  $F_0$ . Left panel : synthetic parameters. Right panel : estimated parameters

Tests were also conducted on natural speech. Two male and two female speakers were asked to produce a fixed vowel, and to vary as much as possible the open quotient. Therefore, they had to press and then to relax their voices, and conversely, to relax and then to press. Acoustic and EGG signals were recorded simultaneously. For a relaxed voice, large values are expected for  $O_q$  and small ones in case of pressed voice. The examples presented here are unstressed to stressed vowels, and then a decrease of  $O_q$  is expected along the vowel. Results for estimation of  $O_q$  for a male and a female speaker are displayed in Figure 6. Open quotient estimation by the spectral method (plain lines) is compared to open quotient measurement using the EGG signal (dotted lines).

As it was expected,  $O_q$  is decreasing. Good agreement between the EGG measurements and the LPC estimation is noticeable. However,  $O_q$  is lower than expected at the beginning, especially for the female. This isn't so surprising, as results on synthetic signals indicated an underestimation of  $O_q$  for higher values.

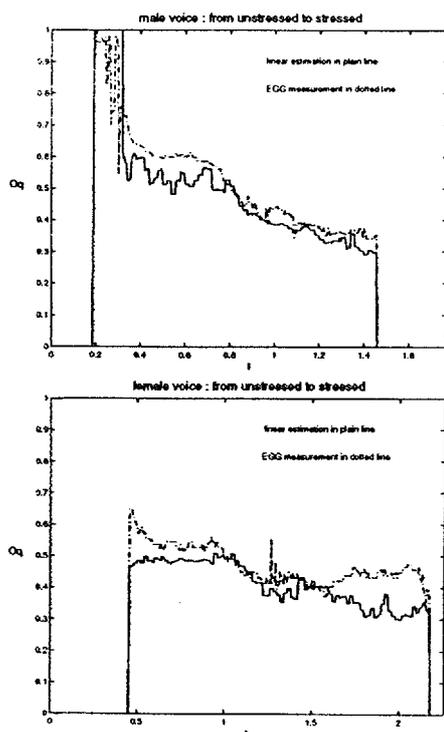


Figure 6: Open quotient estimation for a male (top panel) and a female (bottom panel) voice

## 5 Conclusion

A new method for open quotient estimation is proposed. It is based on a linear model of the glottal flow, and linear predictive analysis of inverse filtered speech. The algorithm performs well for low values of  $O_q$ , but is underestimating high values. The comparison between the estimated  $O_q$  values and the values measured on the corresponding EGG signal showed good agreement.

This study was conducted in the case of an abrupt glottal closure, which is an oversimplified situation. Future work should take this parameter into account.

However we feel entitled to conclude that the spectral method can be applied to voice open quotient estimation in speech and singing.

## References

- [1] Fant G. *Acoustic theory of speech production*. Mouton, La Hague, 1960.
- [2] Miller R.L. Nature of the vocal cord wave. *J. Acous. Soc. Am.*, 31(6):667–677, 1959.
- [3] Sluijter A. van Heuven V. J. and Pacilly J. J. A. Spectral balance as a cue in the perception of linguistic stress. *J. Acous. Soc. Am.*, 101:503–513, 1997.
- [4] Klatt D. and Klatt L. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acous. Soc. Am.*, 87(2):820–857, 1990.
- [5] Childer D.G. and Lee C. K. Vocal quality factors: Analysis, synthesis, and perception. *J. Acous. Soc. Am.*, 90:2394–2410, 1991.
- [6] Alku P. Strik H. and Vilkmann E. Parabolic spectral parameter - a new method for quantification of the glottal flow. *Speech Communication*, 22:67–79, 1997.
- [7] Hanson H.M. Glottal characteristics of female speakers : Acoustic correlates. *J. Acous. Soc. Am.*, 101:466–481, 1997.
- [8] Doval B. and d'Alessandro C. Spectral correlates of glottal waveform models : an analytic study. *Proc. ICASSP'97*, pages 1295–1298.
- [9] Rosenberg A. E. Effect of glottal pulse shape on the quality of natural vowels. *J. Acous. Soc. Am.*, 49:583–590, 1971.
- [10] Fant G. Liljencrants J. and Lin Q. A four-parameter model of glottal flow. *STL-QPSR*, 85(2):1–13, 1985.
- [11] Veldhuis R. A computationally efficient alternative for the liljencrants-fant model and its perceptual evaluation. *J. Acous. Soc. Am.*, 103:566–571, 1998.
- [12] Doval B. and d'Alessandro C. The spectrum of glottal flow models. *notes et documents LIMSI*, 99.
- [13] Markel J.D. and Gray A.II. *Linear prediction of speech*. Springer-Verlag, Berlin, 1976.
- [14] Doval B. d'Alessandro C. and Diard B. Spectral methods for voice source parameter estimation. *Proc. EUROSPEECH'97*, 1:533–536, 1997.