



Spectral correlates of voice open quotient and glottal flow asymmetry: theory, limits and experimental data

Nathalie Henrich, Christophe d'Alessandro, Boris Doval,

LIMSI-CNRS, BP133 - Universit Paris XI, F-91403 Orsay, France

LAM, CNRS - Universit Paris VI - Ministre de la Culture. 11, rue Lourmel F-75015 Paris, France

henrich@limsi.fr, cda@limsi.fr, doval@limsi.fr

Abstract

The effects of voice open quotient and glottal waveform asymmetry are studied in the spectral domain. The hypothesis that the amplitude difference of the first and second harmonics of the inverse-filtered voice signal ($H1^*-H2^*$) is a reliable spectral correlate of the open quotient is tested. Theoretical arguments and experiments are reported. In the theoretical part, analytical formulas are derived for the spectrum of several models (LF, R++, KLGLOTT88). Then it is shown that $H1^*-H2^*$ is generally dependent on both open quotient and asymmetry. Domains for open quotient, asymmetry and $H1^*-H2^*$ variations are given. In the experimental part, examples of voice and singing signals are analyzed. It is shown that a significant part of the spectral measurements obtained are out of the scope of the models studied.

1. Introduction

Studying the spectral correlates of the voice source parameters is of particular value for speech analysis and synthesis (see for instance [2]). This paper is part of an attempt to describe the voice source in the spectral domain using a limited number of parameters [5, 6]. It is often acknowledged in the voice analysis literature that the voice open quotient is closely linked to the relative amplitudes of the first two harmonics. According to Fant/Hanson's terminology [8, 7], Oq is the open quotient retaining the return phase, Oq_i is the open quotient in case of abrupt closure (no return phase). $H1-H2$ is the amplitude difference of the first two harmonics. $H1^*-H2^*$ is the amplitude difference between the two first harmonics, after inverse filtering of the vocal tract (either by actual filtering, or by formant-based correction, like in [7]). $H1^*-H2^*$ is measured on the derivative of the glottal flow spectrum (obtained by some sort of inverse filtering). Hanson [4] stated that "... this discussion has suggested that $H1^*-H2^*$ may be a good measure of OQ". The same assertion can be found in [10], who found a high correlation between $H1^*-H2^*$ measures and OQ. Fant [8] proposed an empiric formulation relating OQ and H^*-H2^* in the framework of the LF-model [1]: $H1^*-H2^* = -6 + 0.27 \exp(5.5Oq_i)$.

In this paper, we present a general set of parameters for glottal flow models, and discuss of their spectral correlates. In the next section, time-domain parameters and spectra are presented. In Section 3, the spectral effects of open quotient and asymmetry is studied, and possible domains of variation for $H1^*-H2^*$ are derived. In section 4, experimental measurements of $H1^*-H2^*$ (speech and singing) are compared to the possible domains predicted by the models. Section 5 discusses of these results and concludes.

2. The spectra of glottal flow models

2.1. Time-domain parameters

Three glottal flow models (GFM) are reviewed in this paper : the LF model [1], the R++ model [9] and the KLGLOTT88 model [3]. Unfortunately, these models do not use the same number of parameters, or the same name for similar parameters. The LF model and the R++ model have 5 parameters. The KLGLOTT88 model is less powerful, with only 4 parameters. We shall show that this difference is significant, particularly for the low and mid-frequencies of the spectrum. A GFM is a function of time, which is always positive or null, and which is periodic, with period T_0 . On a fundamental period, the glottal flow is bell-shaped, with maximum amplitude A_v . It is increasing (opening phase of open phase), then decreasing (closing phase of open phase), then null (closed phase). The open quotient (O_q) is a proportion of the fundamental period (ranging between 0 and 1). It defines also the glottal closure instant (GCI), relative to T_0 (the GCI is at time $t = O_q T_0$). The asymmetry coefficient α_m is the proportion of the opening phase relative to the open phase. It is directly linked to the so-called "speed quotient" (ratio of closing phase /opening phase) by the relation $speedquotient = (1 - \alpha_m)/\alpha_m$. The glottal opening phase is always longer than the glottal closure phase, thus, most models restrict the range of α_m to $[0.5, 0.9]$, with typical values around 0.6 or 0.7. Note that the KLGLOTT88 model has a constant asymmetry coefficient (thus, only 4 parameters). The GFM is a continuous function of time, and is also a differentiable function of time, excepted in some situation at the GCI. In case of "abrupt closure", there is a discontinuity in the glottal flow derivative, resulting in a -12 dB/oct spectral tilt for the GFM spectrum. If this discontinuity is smoothed ("smooth closure", i.e. a differentiable glottal flow at glottal closure), asymptotic spectral tilt is increased. In the time domain, a smooth closure is often described using the so-called "return phase" parameter on the GFM derivative, or alternatively a spectral tilt filter. This "spectral tilt" or return phase coefficient Q_a is the fifth GFM parameter. Another parameter of interest for studying the properties of the glottal flow is the maximum excitation amplitude E of the glottal flow derivative. One can show that E can be computed from the other parameters when necessary. For the LF model (U_g), the correspondance between the original set of parameters and the 5 parameters defined above is as follows:

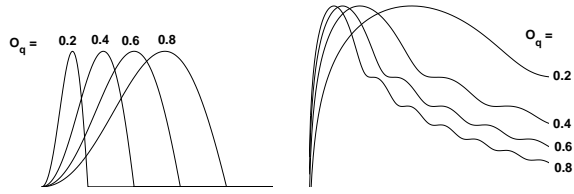


Figure 1: O_q variations shown in time and frequency domains. Left panel: LF model. Right panel: spectra of the derivative of the left panel. F_0 and A_v are arbitrary, $Q_a = 0$, and $\alpha_m = 2/3$.

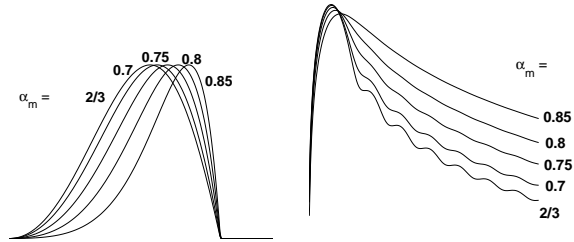


Figure 2: α_m variations shown in time and frequency domains. Left panel: LF model. Right panel: spectra of the derivative of the left panel. F_0 and A_v are arbitrary, $Q_a = 0$, and $\alpha_m = 0.8$.

t_0 fundamental period	$F_0 : F_0 = \frac{1}{t_0}$
t_e open phase	$O_q : O_q = \frac{t_e}{t_0}$
t_p opening phase	$\alpha_m : \alpha_m = \frac{t_p}{t_e}$
t_a return phase	$Q_a = \frac{t_a}{(1-O_q)*t_0}$
E	$A_v = U_g(t_p)$

2.2. Spectra: analytic formulas

Analytic formulas for the spectra of several glottal flow models have been derived in [11]. Because of the limited space available, we shall present herein only the results for the LF and KLGLOTT88 models. After some tedious algebra, one can show that the spectrum of the LF model derivative, for an abrupt closure ($Q_a = 0$), is given by:

$$\tilde{U}'_g(\nu) = A_v \frac{1}{1 + e^{a\alpha_m}} \frac{\pi^2 + \alpha_m^2 a^2}{\pi^2 + \alpha_m^2 (a - 2i\pi O_q T_0 \nu)^2} \left(1 - e^{-2i\pi O_q T_0 \nu} - e^{-a - 2i\pi O_q T_0 \nu} \frac{\sin(\pi/\alpha_m)}{\pi/\alpha_m} 2i\pi O_q T_0 \nu \right)$$

where a must satisfy the implicit equation :

$$e^{-a} + a \frac{\alpha_m}{\pi} \sin\left(\frac{\pi}{\alpha_m}\right) - \cos\left(\frac{\pi}{\alpha_m}\right) = 0$$

The spectrum of the KLGLOTT88 model is given by:

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

With the help of these equations, it is possible to study the spectral features of the GFM. Figures 1 and 2 are showing the effect of O_q and α_m in time and frequency domains. Depending

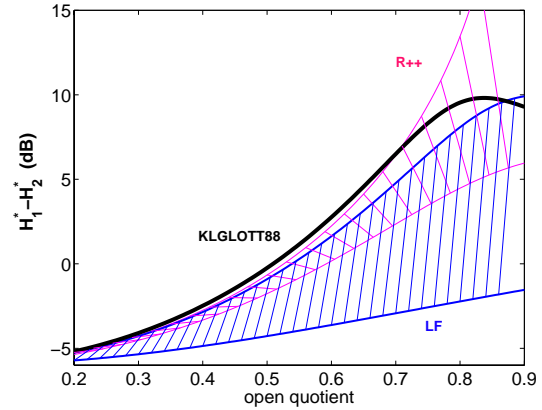


Figure 3: Domain of variation for $H_1^* - H_2^*$ as a function of O_q and α_m , incase of LF, R++ and KLGLOTT88 models ($0.51 \leq \alpha_m \leq 0.75$ for R++ model and $0.66 \leq \alpha_m \leq 0.9$ for LF-model).

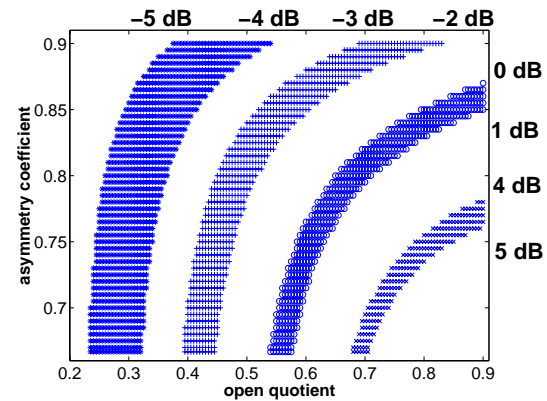


Figure 4: Domain of variation for O_q and α_m for several values of $H_1^* - H_2^*$ ($[-5,-4]$ dB, $[-3,-2]$ dB, $[0,1]$ dB, $[4,5]$ dB) in case of LF model.

on F_0 , both parameters may have an effect in the lower part of the spectrum. In the following, we shall concentrate on the effect of time-domain parameters on $H_1^* - H_2^*$.

3. $H_1^* - H_2^*$ and time-domain parameters

In this section, we shall discuss whether $H_1^* - H_2^*$ is a good spectral correlate of O_q or not. First a particular situation where this statement it actually true is shown. Unfortunately, this is not the general case, and we show that for 5-parameters models, $H_1^* - H_2^*$ depends not only on O_q .

3.1. Open quotient in the KLGLOTT88 Model

In a series of studies, Hanson [7, 4] estimated O_q with measurements of $H_1^* - H_2^*$. Her work is based on the KLGLOTT88 model. In this model [3], it is easy to show that the asymmetry coefficient α_m is a constant value ($\alpha_m = 2/3$). In this case, there is a direct relation between O_q and $H_1^* - H_2^*$. This is shown in Figure 3. However, it must be pointed out that this is the only GFM for which this relation holds, because it is only a 4-parameter model.



3.2. H1*-H2* in the LF and R++ models

Contrary to the KLGLOTT88 model, the LF-model and the R++ model are defined with a set of 5 parameters. In both cases, the additional parameter can be interpreted as an asymmetry parameter α_m (which is kept constant in the KLGLOTT88 model). Then, the relationship between the low frequencies of the spectrum and open quotient is no longer simple. In the case of abrupt closure ($t_a = Q_a = 0$), the asymmetry coefficient must be taken into account. This is demonstrated for instance in Figures 3 and 4. H1*-H2* is plotted as a function of asymmetry and open quotient. These pictures show that for a same value of H1*-H2*, many possible couples of (O_q , α_m) exist. And there is no way to decide which values are correct, given only one spectral measure. Note that it may also be necessary to take into account the effect of the return phase, in case of smooth closure (higher spectral tilt). It seems that the parameter Q_a has indeed an effect on the measure H1*-H2*, but that this effect is much less important than the effect of Oq and asymmetry.

In summary, several combinations of parameters (mainly O_q, α_m , and marginally Q_a) may give exactly the same value for H1*-H2*. Therefore, one can conclude that H1*-H2* can not in general be taken as a reliable measure of O_q . In the general case, this measure provides only a possible domain for O_q and α_m . Additional measures are needed, in order to decide which actual parameter setting corresponds to the data analyzed. If a simpler model is used, without asymmetry, H1*-H2* may be an estimation of O_q . But we know that such a model is not likely to describe the voice source accurately, as the asymmetry parameter is obviously changing for different types of phonation (soft, normal, loud, pressed).

4. Experimental data in speech and singing

In this section, we report some measurements that have been performed on singing signals. Two types of inverse filtering are used. The first one is based on Linear Predictive Analysis (autocorrelation method). The second one is a formant based amplitude correction of the first harmonics of the voiced signal, following the Hanson's formula [4]. The value $20 * \log_{10} \left(\frac{F1^2}{F1^2 - (iF_0)^2} \right)$ is subtracted from the harmonics amplitudes, where F1 is the center frequency of the first formant and iF_0 is the frequency of harmonic i . These measurements are then compared to possible values of H1*-H2*, and it appears that they are out of the scope of all GFM.

4.1. Domain of variation for H1*-H2*

We have seen in the previous section that the link between H1*-H2* and time-domain glottal flow parameters may be intricate. Another problem is the domain of possible variation for this parameter : each specific model allows for a specific domain. It implies that experimental data may, or may not be interpreted in term of each specific model : all the model do not have the same power of explanation for experimental data. In Figure 3 the domain of variation of H1*-H2* for the LF and the R++ models are plotted, as a function of Oq and α_m . Figure 5 displays the H1*-H2* values obtained for a sustained vowel /a/ in speech. The male speaker has been asked to gradually press his voice (sound example 1). This picture shows H1*-H2* measured on the acoustic signal, on LPC-inverse filtered speech, and on formant corrected inverse filtered speech. After careful checking of the data and analysis, it seems that a large part of datas cannot be explained by e.g. the LF model. Open quotient has been

measured directly using an electroglottographic (EGG) reference.

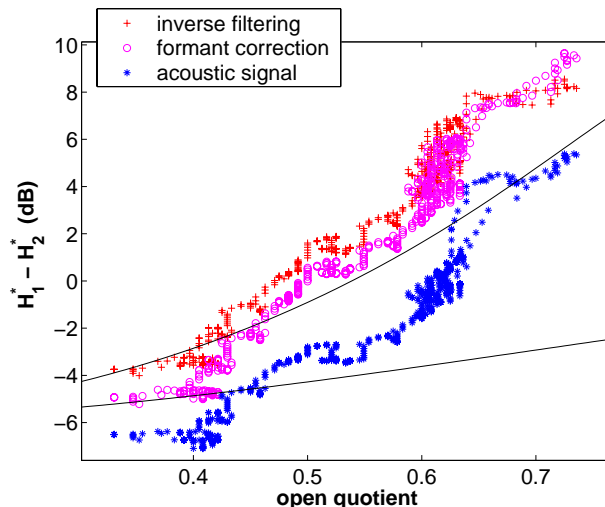


Figure 5: Open quotient measurements. Male voice, sustained vowel /a/. Relaxed (large open quotient) to pressed voice (low open quotient). The solid lines indicate the possible H1*-H2* domain for the LF model.

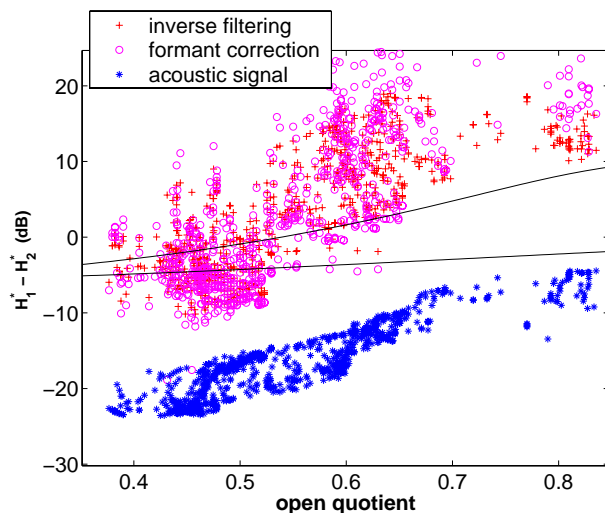


Figure 6: Open quotient measurements. Crescendo on a sustained vowel /a/, C4 ($F_0 = 261 Hz$), sung by a baritone.

4.2. Measurements in singing

A data-base of singing voice has been recorded in studio conditions, and it currently contains various samples of singing for 18 professional singers. Acoustic and electroglottographic signals have been recorded simultaneously [12]. The EGG signal proved useful for accurate and reliable open quotient estimation. Again, spectral estimations of the glottal flow derivative has been performed on inverse-filtered signals, using two different methods. H1*-H2* measures have been obtained directly on these inverse filtered spectra. An example of analysis for a baritone voice is showed in Figure 6 (crescendo on the vowel /a/,



sung at pitch C4 (261 Hz)), sound example 2. In many cases, the $H1^*-H2^*$ measures obtained are compatible with no parameters setting. They can not be explained by any model of the glottal flow signals as they are out of their scope.

5. Discussion and conclusion

The main findings of this research are rather negative. The first point concerns the spectral correlates of open quotient. It has been shown that open quotient can hardly be estimated alone using spectral measures. As a matter of fact, for all the glottal flow models, several different combinations of open quotient and asymmetry may result in exactly the same value for $H1^*-H2^*$. Without further information, it is thus impossible to decide what is the actual value for open quotient. This implies that a more global spectral modeling is needed for parameter estimation: the correspondance between time-domain and spectral parameter is global, and can hardly be concentrated on the first two harmonics.

This problem may go unnoticed with the KLGLOTT88 model, because it uses only 4 parameters: in this case, one can indeed find an explicit relation between O_q and $H1^*-H2^*$. But this parameter estimation is possible only because the model is not as refined as other: it lacks asymmetry, or conversely speed quotient.

However, this result must be tempered. In case of speech or singing production, all parameter combinations are not possible. For instance, a lax voice will have both a large open quotient, high spectral tilt, little asymmetry and low amplitude of voicing. Then, for such a voice, a situation with high asymmetry should be discarded, and only few parameter combinations may be valid.

Sundberg [10] reported a high correlation between open quotient and $H1^*-H2^*$ for barytone voices. This may indicate again that all the parameters are not completely free, but that they are varying jointly.

Another explanation would be that several different glottal parameters configurations may give rise to a same spectrum. This is not against the experience of singers, for instance, that are able to produce a same sound (and therefore a same spectrum) with different glottal configurations. A well-known example is register changes in singing: singers are able to smooth almost perfectly the change of glottal mechanism involved in different registers (e.g. "chest" to "falsetto"), from the auditory point of view. No or almost no changes in sound are perceived, although the glottal settings are changing drastically.

The second point is even more fundamental. It concern the modelling capabilities of the current glottal flow models. Using inverse filtering of singing signals gave rise to spectra than can not be matched by any glottal flow model. The figures obtained for $H1-H2$ can not be produced, whatever the model and its parameter setting. Note that inverse filtering is based on the underlying source/filter hypothesis, i.e. independance of the source and filter components of voice production. This is of course only an approximation, and in case of strong voice/filter interaction, inverse filtering may no more be valid.

In summary, this negative result may be interpreted along two lines: on the one hand it may indicate that current models are not able to deal with all the variability of vocal production; on the other hand, it may indicate that source/filter interaction is large enough to hinder a valid inverse filtering approach.

These results raise the question of auditory perception for glottal source parameters. A high precision is not required for parameter estimation if parametric variations go unnoticed for

listeners. Future work will be devoted to time-domain parameters and spectral parameters perception for glottal source signals.

Sound examples for this paper can be found at : <http://www.limsi.fr/Individu/henrich/Eurospeech01.html>

6. References

- [1] Fant, G. Liljencrants, J. and Lin, Q. (1985) "A four-parameter model of glottal flow", *STL-QPSR*, **85(2)**, 1-13.
- [2] Fant, G. and Lin, Q. (1988), "Frequency domain interpretation and derivation of glottal flow parameters." *STL-QPSR*, **88(2-3)**, 1-21.
- [3] Klatt D. and Klatt L. (1990) "Analysis, synthesis, and perception of voice quality variations among female and male talkers" *J. Acoust. Soc. Am.* **87**, 820-857.
- [4] Hanson, H. M. (1995) *Glottal characteristics of female speakers* (Ph.D. Thesis, Harvard Univ.).
- [5] Doval, B and d'Alessandro, C. (1997). "Spectral correlates of glottal waveform models: an analytic study." in *Proc. ICASSP 97, Munich*, pp 446-452.
- [6] Doval, B. d'Alessandro, C. Diard, B. (1997) "Spectral methods for voice source parameter estimation." *Proc. EUROSpeech'97, Vol. 1*, 533-536.
- [7] Hanson, H. M. (1997) "Glottal characteristics of female speakers: Acoustic correlates" *J. Acoust. Soc. Am.* **101**, 466-481.
- [8] Fant, G. (1997), "The voice source in connected speech." *Speech Communication*, **22**, 125-139.
- [9] Veldhuis, R. (1998). "A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation" *J. Acoust. Soc. Am.* **103**, 566-571.
- [10] Sundberg Johan, Andersson Maria and Hultqvist Clara (1999) "Effects of subglottal pressure variation on professional baritone singers' voice sources" *J. Acoust. Soc. Am.* **105**, 1965-1971.
- [11] Doval, B., d'Alessandro, C. (1999) "The spectrum of glottal flow models", LIMSIS report, NDL N 99-07, submitted for publication, 22pp.
- [12] Henrich, N., Doval, B., d'Alessandro, C., Castellengo, M. (2000) "Open quotient measurements on EGG, speech and singing signals" 4th workshop "Advances in Objective Laryngoscopy, Voice and Speech Research", Jena, Germany, April 7-8, 2000.