ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 92 (2006) 1 – 1

The Spectrum of Glottal Flow Models

Boris Doval, Christophe d'Alessandro
LIMSI - CNRS, BP 133 F-91403 Orsay, France
Nathalie Henrich
ICP - INPG, 46 Avenue Félix Viallet F-38031 GRENOBLE, France

Summary
A unified description of the most-common glottal-flow models (KLGLOTT88, Rosenberg C, R++, LF) is proposed in the time domain, using a set of five generic glottal-flow parameters: fundamental period, maximum excitation, open quotient, asymmetry coefficient, and return-phase quotient. A unified set of time-domain equations is derived, and their analytical Laplace-transform computation leads to a set of frequency-domain equations. On the basis of this mathematical framework, the spectral properties of the glottal-flow models and their derivatives are studied. It is shown that any glottal-flow model can be described by a combination of low-pass filters, the cut-off frequencies and amplitudes of which can be expressed directly in terms of time-domain parameters. The spectral correlates of time-domain glottal-flow parameters are then explored. It is shown that the maximum excitation corresponds to a gain factor, and that it controls the mid-to-high-frequency spectral slope. A non-null return-phase quotient adds an additional spectral tilt in the high-frequency part of the glottal-flow spectrum. The open quotient and asymmetry coefficient are related to the low-frequency spectral peak, also called the glottal formant. The glottal-formant frequency is mainly controlled by the open quotient, and its amplitude (or bandwidth) by the asymmetry coefficient. As a direct application, it is shown that the amplitude difference between the first two harmonics, commonly assumed to be correlated to the open quotient, is also theoretically dependent on the asymmetry coefficient.

PACS no. 43.70.Bk, 43.70.Gr, 43.72.Ar

1. Introduction

Linear acoustic theory describes speech production in terms of a source/filter model [1]. This model consists of a volume velocity source, which represents the glottal signal, a filter, associated to the vocal tract, and a radiation component, which relates the volume velocity at the lips to the radiated pressure in the far acoustic field. From the point of view of physics, this model is only an approximation, whose main advantage is simplicity. It is considered valid for frequencies below 4 to 5 kHz, where the assumption of plane wave propagation in the vocal tract seems acceptable.

Most of the speech features related to voice quality, vocal effort, and prosodic variations can be associated with the voice source function. Thus, modelling this component of the speech production model is essential in speech analysis and synthesis, and speech perception. The approach that is chosen in this study is voice source signal modelling, following the pioneering work of Fant [2], Rosenberg [3] and others. In this signal analysis approach, most of the glottal flow models proposed are time-domain models [3, 4] (KLGLOTT88 model), [2] (LF model), [5] (R++ model), [6]). Time-domain models have a number of advantages. The first is that timing relationships are very important for modelling the glottal flow signal. The model parameters are always functions of some temporal phases of the signal, like the fundamental period, the open quotient (a measure of the relative open duration) or the speed quotient (a measure of the flow asymmetry). Another reason for studying the glottal flow in the time domain is that the glottal activity can be studied along with other time-domain physiological analyses, like electroglottography, high-speed cinematography or electromyography.

In areas such as speech synthesis, voice quality analysis, or speech processing, a frequency-domain approach also appears to be desirable. Generally, voice quality is better described by spectral parameters. For instance, Hanson [7] or Klatt [4] found that the main spectral parameters 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. Several authors were interested in measuring the rate of decay of the voice source spectrum. Childers and Lee [8] presented the harmonic richness factor for measuring the source spectral decay. This parameter is the amplitude ratio between the fundamental and the sum of the higher harmonics. Alku *et al.* [9] presented the parabolic spectral parameter. This parameter represents the rate of decay of the low frequencies, for the inverse-filtered pitch-synchronous spectrum. It may also be related to some features of differ-

Received 1 January 2005,
accepted 1 January 2005.

ent phonation types. Oliveira [10] used an analytical formula for the spectrum of the KLGLOTT88 model. He estimated the model parameters in the frequency domain by minimization of the differences between speech parameters and the model parameters. Fant and Lin [11] used some approximations of the LF model spectrum rather than any analytical formulation. They re-analyzed an old database in order to estimate source parameters. Doval *et al.* [12] focused on the open quotient estimation by spectral fitting of a 2nd order model of inverse filtered speech, according to an analytical formulation of the KLGLOTT88 model. In the present situation, the exact correspondence between time-domain parameters and the spectrum is still unclear, mainly because most work has been empirical, without analytical formulas for the spectrum of those models, and without a spectral model of the glottal flow.

In most of the studies we have been able to locate on the spectrum of glottal flow, the spectrum has been obtained by numerical Fourier transform of the glottal waveform. Therefore, it seemed important to develop analytical formulae for the spectrum of glottal flow, in order to gain better insight into the spectral features and properties of the voice source. Moreover, in some applications, like speech analysis/synthesis, where it would be important to change voice quality, a spectral model of the glottal flow signal would be desirable. It must be emphasized that the spectral domain is mathematically equivalent to the time domain, only if complex spectra are considered. In this case, time and frequency domains are related through the Fourier transform. But spectral magnitudes and spectral phases do not play the same role (although they are merged in time domain). For speech processing, one can take advantage of this separation in the frequency domain because spectral processing does not require the use of calibrated measurement equipment for speech recording. For example, phase distortion is acceptable for spectral processing, although it is a well-known source of problems for time-domain processing, because it can significantly change the signal waveform.

It is acknowledged that the general shape of the glottal flow magnitude spectrum is low-pass. Thus, the magnitude spectrum of the glottal flow derivative is band-pass, with a single maximum located in the vicinity of the fundamental frequency (F_0) or its second harmonic. This spectral maximum has been called the "glottal formant" [13], because it appears in the speech spectrum as a spectral maximum, generally located below the first vocal tract formant (F_1). The term "voicing bar" is sometimes used for this first spectral maximum observed on spectrograms. The aim of this work is to study the position, variation and properties of the glottal formant in a unified framework, and to give explicit equations for relating time-domain glottal flow parameters to the glottal formant.

This study has many interesting outcomes: 1. it defines the spectral behaviour of most common glottal flow models and, more specifically, the variation of the voice source in the frequency domain. A set of spectral parameters is proposed. 2. it gives the relationships between time-domain and spectral parameters. 3. it gives some hints for spectral estimation of glottal flow parameters and spectral modification of glottal flow parameters.

In a first part of this paper (section 2), a unified view of time-domain glottal flow models is proposed. It is shown that the KLGLOTT88, Rosenberg C, R++ and LF models, and all the time-domain models built on similar principles, can be represented using a unified set of parameters. It is then possible to consider the properties of all these models in a common framework.

The glottal formant is then introduced in section 3. It is shown that the main features of glottal flow spectra can be represented by the frequency response of a combination of low-pass filters (defined by the "glottal spectral peak", and the "spectral tilt filter"). The links between parameters of the glottal flow spectra and time-domain parameters of the glottal flow waveform are established.

In section 4, the variation of the glottal formant is computed. These results give some insights into the possible variations of the glottal formant, and into the relationships between variations of time-domain parameters and frequency domain parameters. The relationships between harmonic amplitudes, open quotient and other parameters are specifically studied. Then the spectral correlates of time-domain glottal flow parameters are reviewed. Sound examples accompany these analyses, in order to enhance the reader's understanding of the spectral variations of glottal flow models. A summary of the results obtained is given in section 5.

2. Time-domain description of glottal flow models

The aim of this section is to analyze in a common framework some glottal flow models (GFM) that have been proposed in the literature: KLGLOTT88 [4], Rosenberg C [3], R++ [5] and LF [2] models. The last of these is the most used, but the others have been included for the sake of generality.

First, it is shown that all these models have essentially the same common features, despite different presentations and different sets of parameters. A common set of 5 generic time-domain parameters is proposed for these models. It is then possible to derive a unified description that allows the specific features of each model to be understood, using the generic parameters and a generic form of each model. This unified GFM description will be useful for computing their spectra in section 3, and for studying spectral and time domain correspondences in section 4.

2.1. Time-domain glottal flow features

Several glottal flow models have been proposed so far. However, they generally do not use the same number of parameters, or the same name for similar parameters. Initially this makes it rather difficult to understand the similarities and differences between models. In this section a common framework for analysis and representation of all

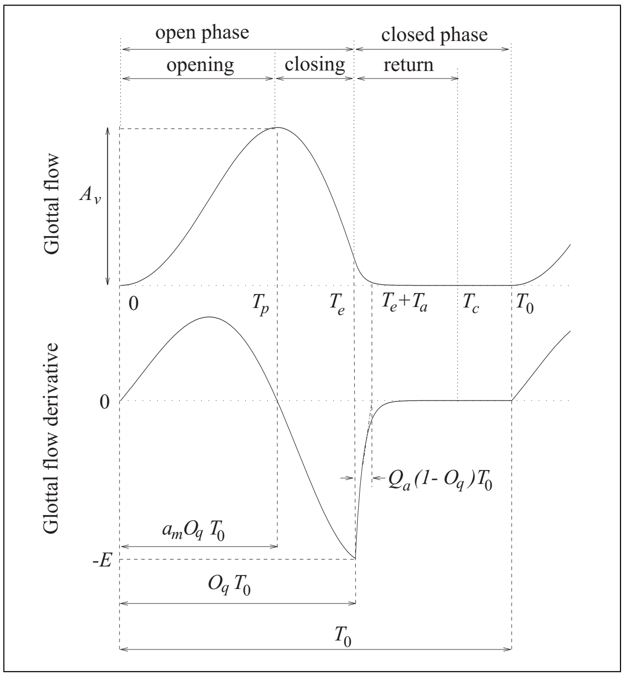

Figure 1. Phases and parameters of the glottal flow and its derivative. See section 2. (a LF model has been used to obtain the curves).

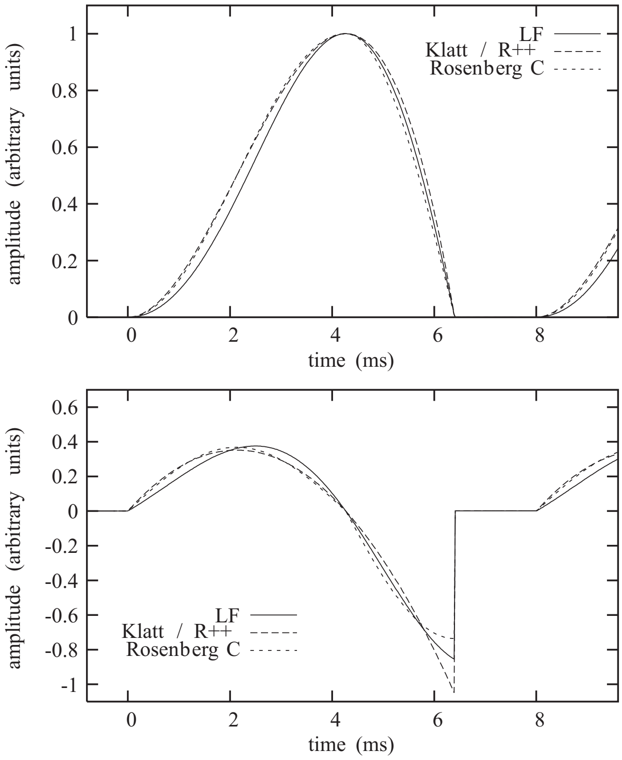

Figure 2. Example of 4 GFMs (top) and their derivatives (bottom) with abrupt closure and with a common set of parameters: $T_0 = 8ms$, $O_q = 0.8$, $\alpha_m = 2/3$ and $A_v = 1$. KLGLOTT88 and R++ models are identical for this parameter set. The waveforms are very similar. Note that E differs between models when A_v and the other parameters are fixed.

these models is presented. All the models share the following common features:

- the glottal flow is always positive or null.
- the glottal flow is quasi-periodic.
- during a fundamental period, the glottal flow is bell-shaped: it increases, then it decreases, then it becomes null.
- the glottal flow is a continuous function of time.
- the glottal flow is a differentiable function of time, except in some situations at the instant of glottal closing.

The glottal-flow derivative is often considered in place of the glottal flow itself. This is because, in the voice production model, the transfer function of the radiation component, relating the acoustic flow at the lips to the acoustic pressure in the acoustic field in front of the lips, can be considered as a derivative to a first approximation. Also, the shape of the glottal flow derivative can often be recognized in acoustic speech or singing voice waveform itself. For instance, the peak of the derivative is often visible. An example of GFM and GFM derivatives, and the GFM derivative with a vowel, is given in sound examples 1 and 2.

The glottal-flow derivative for all the models shares the following common features:

- the glottal-flow derivative is quasi-periodic.
- during a fundamental period, the glottal-flow derivative is positive (when the glottal flow is increasing), then null (glottal flow maximum), then negative (when the glottal flow is decreasing), then null (when the glottal flow is null).
- the glottal flow derivative is a continuous function of time, except in some situations at the instant of glottal closing.
- the glottal flow derivative is a differentiable function of time, except at the instant of glottal closing.

Generally, all the glottal flow models are described in terms of phases in the time domain. Let us consider these phases during one fundamental period T_0 (Figure 1).

The first phase is the opening phase, when the glottal flow increases from baseline at time 0 to its maximum amplitude. The maximum amplitude of the glottal flow is the so-called amplitude of voicing A_v and is reached at time T_p.

The second phase is the closing phase, when the flow decreases from A_v to a point at time T_e where the derivative reaches its negative minimum. This time T_e is the glottal closing instant (GCI). The absolute amplitude of this minimum is called the maximum vocal-tract excitation (or maximum excitation) E. Then the GCI corresponds to the time of maximum speed of vocal-fold closure. In studies with real speech, the glottal closing phase is often treated as though based on the flow (and not on the flow derivative) and it is measured from the instant of the flow maximum to the instant of closure, which is estimated as a time instant when the flow comes back to the DC level. As defined above, the present study uses a slightly different

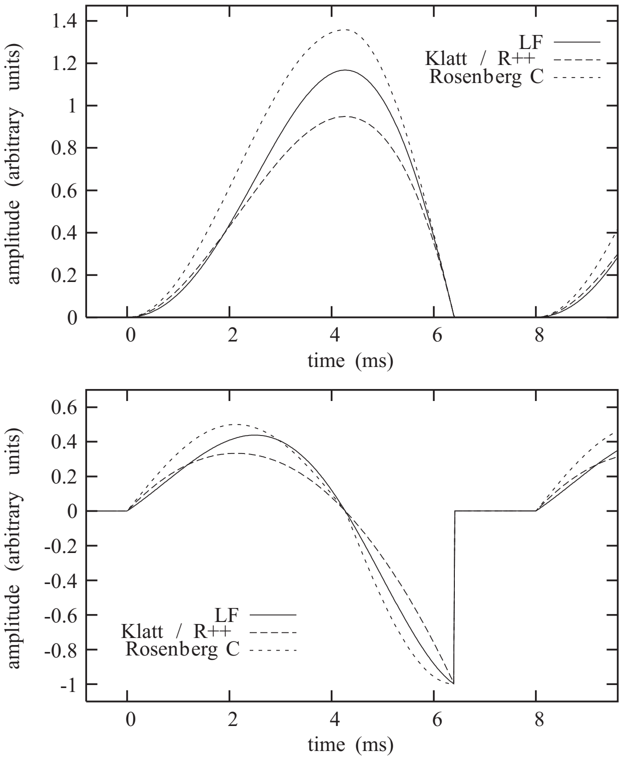

Figure 3. Same Figure as Figure 2 but when E is kept constant ($E = 1$). Note that A_v differs between models when E and the other parameters are fixed. However these differences are hardly audible (sound examples 3, 4 and 5).

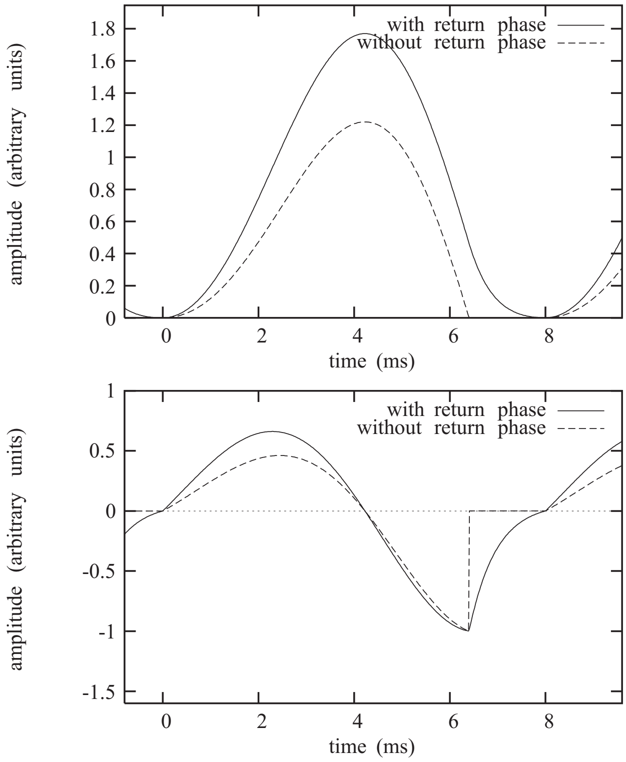

Figure 4. GFM (top) and its derivative (bottom) with abrupt or smooth closure. Note that the return phase smoothes the discontinuity at GCI but also slightly changes the whole waveform shape.

notation, where the closing phase ends at the GCI when the derivative reaches its minimum value.

These two phases (opening and closing phases) result in the open phase, which is characterized by the open quotient O_q. This quotient is defined as a fraction of T_0, thus ranging from 0 to 1: $O_q = T_e/T_0$. The ratio between the opening phase and the closing phase is the speed quotient $S_q = T_p/(T_e - T_p)$.

The last phase is the closed phase. Two main cases must be distinguished according to the shape of the GFM at the GCI. The first case corresponds to “abrupt closure”. In this case, there is a discontinuity in the glottal flow derivative which instantaneously reaches 0 after maximum excitation. Then the left and right derivatives of the glottal flow exist but are different at the GCI. Mathematically, this corresponds to a non-differentiable glottal flow at this point. The glottal flow (and its derivative) are null between O_qT_0 and T_0. Examples of GFM with an abrupt closure are displayed in Figures 2 and 3.

The second case corresponds to “smooth closure”. Mathematically the glottal flow is differentiable at the GCI, as the derivative is continuous. In the time domain, a smooth closure introduces a “return phase” which takes place between the GCI and the closure instant T_c. Frequently, the time T_c is taken as T_0, which means that the glottal flow and its derivative will never be null in the closed phase except at the single point T_c. The general form of differentiable GFM is displayed in Figure 4. Mathematically the different models use one of two ways to smooth the derivative at the GCI: either adding a decreasing exponential (“return phase method”) or passing it through a low-pass filter (“low-pass filter method”). The return phase method has been originally proposed by Fant [2] in the LF model. Veldhuis [5] proposed a general approach for using the return phase method for any time-domain GFM. It consists in adding a decreasing exponential to the GFM derivative between the GCI and the closure instant. This exponential is parameterized by its time constant T_a which characterizes the speed of return. Because the exponential always has to be decreasing, this value cannot be greater than the duration between the GCI and the end of the period: $T_a < T_0 - T_e$. The drawback of this method is that it implies the use of a supplementary hidden parameter, the value of which has to be obtained by resolving an implicit equation. This is due to the constraint that the GFM derivative must have a null integral, so that the GFM must have the same value at its start and end points (the baseline is constant). However, to avoid this drawback, some authors have proposed a floating GFM baseline, achieved by modelling the return phase independently of the first phases [14].

The low-pass filter method has been introduced by [4]. It consists of filtering the GFM derivative by a first-order (or second-order) low-pass filter. It mainly affects the discontinuity at the GCI, convolving it by the filter impulse response. As this impulse response is also a decreasing exponential, it can be parameterized by its time constant T_a, and the corresponding return phase is very close to this exponential.

It should be noticed that, except in the case where no constraint is applied on the baseline for the "return phase" method, both methods modify the shape of the opening and closing phases even if the other parameters are kept constant. However the position of the GCI is not changed. But while the return phase method does not change the position of the glottal flow maximum, the low-pass filter method not only changes it but introduces non null values after the end of the period, since the impulse response of the filter is infinite.

To take the return phase into account, Fant and others proposed a modified definition of the open quotient as: $O_q = (T_e + T_a)/T_0$. However, in this paper, we do not use this modified definition.

2.2. A common set of time-domain parameters

The different GFMs are described by different sets of parameters. But, as a matter of fact, many alternative sets of parameters can be used to describe a single model. For instance, for a LF model, the parameters (T_e, T_p, T_a) are often replaced by the parameters (R_g, R_k, R_a) [15]. And fortunately the same parameters can very often be used for different models with the same meaning (think of the open quotient, for instance). Thus it is necessary to search for a common set of parameters to describe the studied GFMs, even if this common set implies a rewrite of the relevant mathematical expressions.

One way to obtain such a common set of parameters is to consider each GFM as a family of curves which is specified by control points through which the curves must pass (there will be as many control points as the number of parameters). If we define these points by the general properties of the glottal flow expressed above, then the control points will be independent of a particular GFM. For example, $(T_p,0)$ or $(T_e,-E)$ are control-point candidates of the GFM derivative. In this framework, we chose as control points of the GFM derivative: (0,0), $(T_p,0)$, $(T_e,-E)$, $(T_0,0)$ and the right-side derivative of point $(T_e,-E)$ to specify the return phase.

We are now left with the choice of a parameter set that describes these control points. For instance (E, T_0, T_e, T_p, T_a) and (E, T_0, O_q, S_q, T_a/T_0) are possible sets. A good parameter property is to be dimensionless. In particular this allows the bounds for useful parameter values not to depend on other parameters. For example, useful values for the open quotient are from 0.3 to 0.8, but the corresponding bounds on T_e depend on T_0 ($0.3 * T_0$ to $0.8 * T_0$). The same is true for the speed quotient which generally lies between 1.5 and 4, while the corresponding time T_p depends on T_e which itself depends on T_0.

Therefore, from now on, we shall use the following set of 5 time-domain parameters:

- E, the maximum excitation.
- T_0, the fundamental period.
- O_q, the open quotient, defined by the ratio between the open phase duration and the fundamental period.
- α_m, the asymmetry coefficient, defined by the ratio between the opening phase and the open phase durations.
- Q_a, the return phase quotient, defined by the ratio between the return phase time constant and the duration between the GCI and the end of the period: $Q_a = T_a/[(1 - O_q)T_0]$.

The parameter E generally corresponds quite well with the main speech waveform peak in the time domain. Therefore, in practice, it may be more appropriate than the amplitude of voicing A_v.

The parameter O_q controls the relative duration of the glottal flow pulse as it defines the GCI relatively to T_0: the GCI is at time $T_e = O_q T_0$. Thus, if O_q is small, the glottal pulse is narrow; if O_q is larger, the glottal pulse is wider. More precisely, the flow is stretched proportionally to Oq. O_q ranges from 0 to 1. It is related to pressed versus relaxed voice quality [4]. This parameter can be measured during phonation using electroglottography [16] without the need for any glottal flow model.

The parameter α_m controls the degree of asymmetry of the glottal pulse as it defines the instant of maximum of the glottal flow relatively to T_0 and O_q: this instant is at time $T_p = \alpha_m O_q T_0$. Theoretically it ranges between 0 and 1 but, for real speech Signals, the glottal opening phase is always longer than the glottal closing phase (see [17] where Titze has shown with the help of a physical model that this skewing to the right is mainly attributable to source-tract nonlinear interaction). Thus, α_m range is restricted to $[0.5, 1]$, with typical values around 0.6 to 0.8. For mathematical reasons, most models restrict this range again ($[0.65, 1.0[$ for LF, $]0.5, 0.75]$ for R++, see details in Appendix A2). When α_m is high, the flow is very asymmetric and it has an impulse-like shape. When α_m is low (near 0.5), the flow tends towards a sinusoidal waveform. The asymmetry coefficient, which has been introduced by [18], can be related to the speed quotient by: $\alpha_m = S_q/(1 + S_q)$.

The return phase quotient, introduced in [19], is the relative duration of the return phase. It ranges from 0 to 1, the special case $Q_a = 0$ corresponding to an abrupt closure. When Q_a is high, the glottal flow discontinuity is smoothed, corresponding to soft voices.

Another interesting glottal flow parameter, which is not described in the literature, is the total flow I, defined by the glottal flow integral between 0 and T_0. Its importance will become apparent when the spectrum is described. This parameter can be deduced from the 5 preceding parameters as we will see in the following section.

2.3. A generic glottal flow model

All the glottal flow models can be described using the set of 5 parameters defined above, thanks to some straightforward algebra, as is shown in Appendix A1 for the four models considered. This raises the question of the differences between models: what makes a particular model unique? To better answer this question, we would like to concentrate on how each model performs with our 5 parameters. The amplitude and the periodicity parameters are easily treated, since their effects are independent of any given model. The return phase parameter effect will be treated separately (see section 3.4). We are left with the

2 last parameters O_q and α_m. It is show below that, in the case of abrupt closure and for all the models we have studied, O_q can be factorized in the expression for the GFM. In other words, its effect (shrinking or stretching the time scale) will be the same whatever the model.

To go into more detail, we must first consider $P = (E, T_0, O_q, \alpha_m, Q_a)$, the set of time-domain parameters that have been chosen and $U_g(t; P)$, the expression of the glottal flow which is a function of these parameters. The derivation is easy in the case of an abrupt closure ($Q_a = 0$). In this case, each parameter can be isolated in the mathematical expression of the glottal flow U_g. The glottal flow considered in only one period can be rewritten as follows, whatever the model:

$$U_g(t; P) = EO_qT_0n_g\left(\frac{t}{O_qT_0}; \alpha_m\right), \; 0 \le t \le T_0, \quad (1)$$

where $n_g(\tau; \alpha_m)$ is a function of time τ which depends on only one parameter, namely α_m, and on which uniqueness of a given model is concentrated. In this paper this function will be called the "generic model". It is obtained by setting $E = 1$, $T_0 = 1$ and $O_q = 1$ in the expression of $U_g(t; P)$. The expression for this generic model is derived for the four models in Appendix A2. The differences between generic models are rather minimal and only lie in the particular mathematical functions used in each case.

As the voice source is periodic, the expression of the glottal flow must be periodically repeated at the rate $F_0 = 1/T_0$:

$$U_g(t; P) = EO_qT_0n_g\left(\frac{t}{O_qT_0}; \alpha_m\right) * \text{Ш}_{T_0}(t), \quad (2)$$

where $\text{Ш}_{T_0}(t)$ is a Dirac comb with fundamental period T_0.

According to equation (2), U_g can be obtained by taking the n_g function between 0 and 1, applying to it a time scaling of O_qT_0 (the function is now defined between 0 and O_qT_0) and an amplitude scaling of E, and finally repeating it at a rate of T_0. Note that O_qT_0 represents a multiplicative coefficient which compensates for the time scaling, so that the final maximum excitation is actually E. Thus E behaves as a gain parameter, O_q and T_0 are time stretching/shrinking parameters, and T_0 is used for periodic repetition. Parameter α_m gives the shape of the glottal flow. The glottal waveform is model dependent according to the n_g expression. However, as can be seen on Figure 2, all the GFMs will look very similar when the same parameters are used.

The amplitude of voicing and the total flow depend on the 4 basic parameters. It is possible to derive these relationships using equation (2). Since the amplitude of voicing A_v is the maximum of U_g in one period, by taking the maximum on both sides of equation 2, it is found that A_v is related to the maximum of n_g, $a_v(\alpha_m)$ according to:

$$A_v = a_v(\alpha_m)EO_qT_0.$$

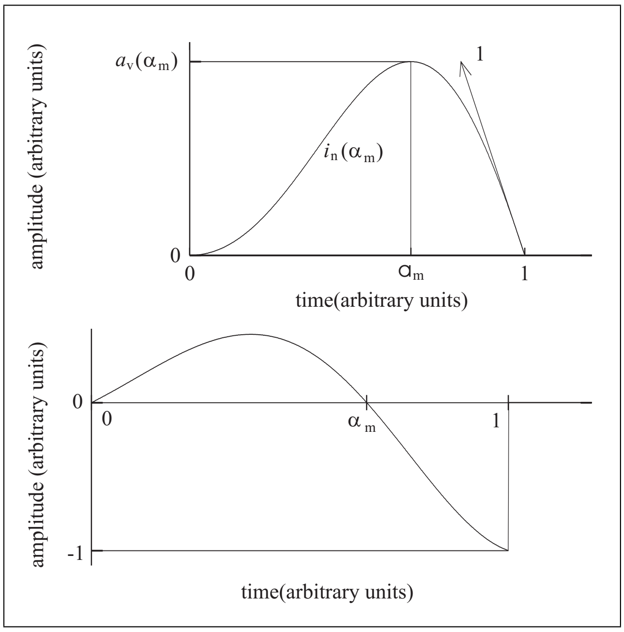

Figure 5. Generic model and its derivative. In the case of abrupt closure, a GFM can be easily computed from its generic model using only the parameter α_m.

Because U_g reaches its maximum at time $T_p = \alpha_m O_q T_0$, by definition, we can also obtain the same result by substituting T_p in equation (2). This leads to a method of computing $a_v(\alpha_m)$: $a_v(\alpha_m) = n_g(\alpha_m; \alpha_m)$ (see Figure 5).

Similarly, since the total flow I is the integral of the glottal flow over one period, taking the integral of both sides of equation 2 shows that I is related to the integral of n_g, $i_n(\alpha_m)$, by:

$$I = i_n(\alpha_m)E(O_qT_0)^2. \quad (3)$$

It is straightforward to obtain the GFM derivatives by taking the time derivative of equation (2):

$$U_g'(t; P) = En_g'\left(\frac{t}{O_qT_0}; \alpha_m\right) * \text{Ш}_{T_0}(t). \quad (4)$$

Unfortunately, in the general case of differentiable GFM (with smooth closure, $Q_a \neq 0$) it is no longer possible to compute analytic expressions for the generic models that would be independent of O_q and T_0. This is because the GFM is no longer null between O_qT_0 and T_0 due to the return phase, and then there is a bound on the definition of n_g that depends on O_q and T_0.

In summary, one can consider that the different glottal flow models are all very similar. All the models can be represented by the same set of 5 parameters. Notice however that the KLGLOTT88 and the Rosenberg C models have only 4 parameters, the asymmetry coefficient being fixed at 2/3 for the KLGLOTT88 model, and the return phase quotient being null for the C model. All the models are pulse-like and the amplitude of the pulse depends on E. The relative width of the pulse depends on O_q. The rate of repetition of the pulses depends on T_0, which is therefore responsible for the voice melody. The pulse shape depends on the asymmetry coefficient α_m in a way that is specific to each model. This asymmetry coefficient mainly controls the skewness of the pulse. The return phase quotient smoothes the glottal waveform at GCI. If all these parameters are fixed, the only remaining difference between models is therefore the specific mathematical functions used for their definition. This difference does not seem to be very significant in practice (see sound examples 3, 4 and 5).

The main advantage of computing a generic model is that it makes clear the similarities of models through a common set of parameters. But the generic model is also useful for computing the spectra and for interpreting the spectral effects of the parameters, as will be shown in the next section. Finally, the generic model of any other GFM may also be computed using the same framework.

3. Spectral description of glottal flow models

In this section, decomposition of glottal flow waveforms using generic parameters and their generic forms is used for deriving the spectrum of glottal flow models. It is shown that GFM can be considered as low-pass filter impulse responses. The frequency responses of four GFM are derived in analytical form. It is then possible to study the main features of these frequency responses and to propose a stylized form of the GFM spectra based on the glottal formant and the spectral tilt. This representation will be used in section 4 for studying the spectral effect of generic time-domain parameters.

3.1. Glottal flow models as Low-pass filters

Since the early years of the source-filter theory of speech production, it is well known that the effect of the glottis in the spectral domain can be approximated by a low-pass system. With this the glottal flow signal is considered as the output of this low-pass system to an impulse train. In a transmission line analog, Fant [1] used four poles on the negative real axis in the form:

$$U_g(s) = \frac{U_{g0}}{\prod_{i=1}^{4}(1 - s/s_{ri})}, \tag{5}$$

with $|s_{r1}| \simeq |s_{r2}| = 2\pi 100\,Hz$, and $|s_{r3}| = 2\pi 2000\,Hz$, $|s_{r4}| = 2\pi 4000\,Hz$. This is a 6 parameters spectral model (F_0, U_{g0} with four poles), 2 parameters being fixed (s_{r3} and s_{r4}). According to Fant, s_{r1} and s_{r2} account for the variability with regard to speaker and stress.

This simple form has had great success, because it has been used for deriving the linear prediction equations (see, for instance [20]). In this latter case, only two poles are used, because the linearity of this acoustic model only holds for frequencies below about 4000 Hz. Such a simple filter depends on only 3 parameters: a gain factor U_{g0}, the fundamental frequency F_0, and a frequency parameter $s_{r1} \simeq s_{r2}$. This spectrum has an asymptotic behaviour in -12 dB/oct when the frequency tends towards

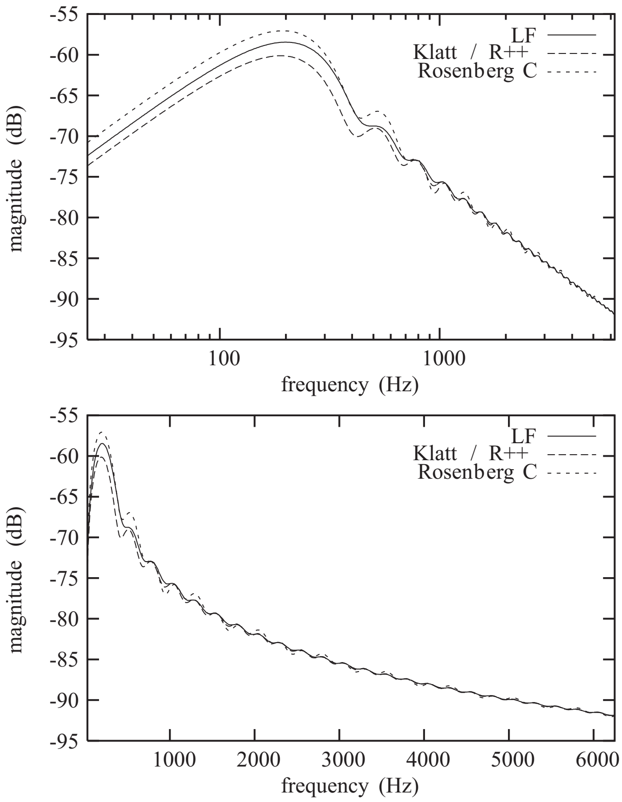

Figure 6. Spectra of the 4 GFM derivatives of Figure 3. The log-frequency scale allows the asymptotes to be more clearly seen.

infinity. The parameter s_{r1} controls the cut-off frequency β of the spectrum. When the frequency tends towards 0, $|U_g(0)| \sim U_{g0}$ and when the frequency tends towards infinity $|U_g(f)| \sim U_{g0}\beta^2/f^2$ where $\beta = s_{r1}/2\pi$. Therefore the spectral tilt is null for frequencies below β, and is of -12 dB/oct for frequencies above β.

This cut-off frequency β remains the same whether it is computed using the GFM or its derivative. This is important because the speech sound is a pressure signal in which the effect of the flow is manifested by its derivative, due to the lip radiation effect. The asymptotic lines have $+6$ dB/oct and -6 dB/oct slopes for the derivative (because of the factor f in the spectrum). The spectral characteristics of the glottal pulse are those of a second-order filter frequency-response, showing a spectral peak near the asymptotic lines crossing point which is then called the glottal formant. This glottal formant is generally noticeable on spectrograms, especially for male voices. It is sometimes referred to as the "voicing bar" in spectrogram reading.

3.2. Spectrum of the generic glottal flow model

The spectrum of the general form of any GFM presented above also corresponds to a low-pass system frequency response, as shown in Figure 6 and 7. In the case of abrupt closure, taking the Fourier transform of equations (2) and (4) gives the spectra $\widetilde{U_g}$ and $\widetilde{U_g'}$ of the GFM and its derivative:

$$\widetilde{U_g}(f;P) = E(O_qT_0)^2\widetilde{n_g}(fO_qT_0;\alpha_m)\big(F_0\text{Ш}_{F_0}(f)\big), \tag{6}$$

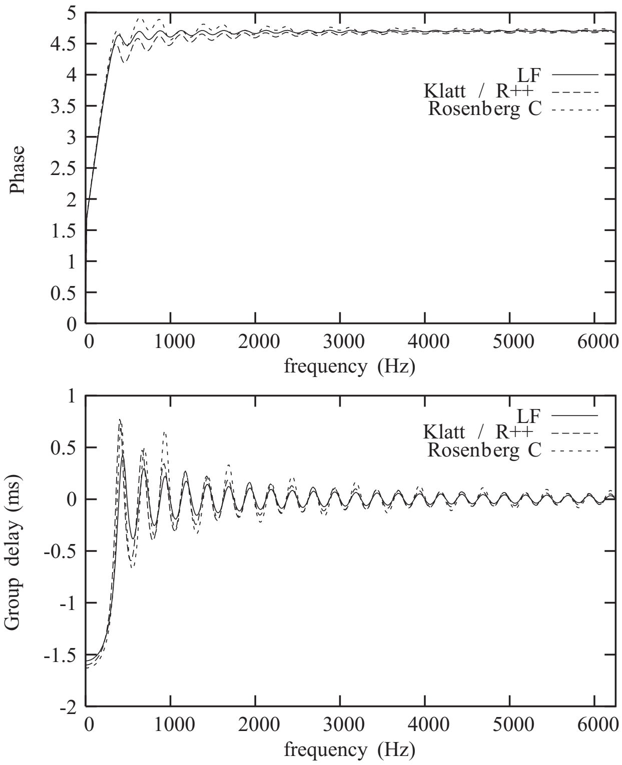

Figure 7. Phase spectra and group delay of the 4 GFM derivatives of Figure 3. Above the glottal formant frequency, the phase is almost constant. The oscillations on the phase and the group delay are due to the finite duration of the GFM. At the frequency of the glottal formant the group delay is negative, which shows that this maximum is anticausal.

$$\widetilde{U'_g}(f;P) = EO_qT_0\widetilde{n'_g}(fO_qT_0;\alpha_m)\big(F_0 ш_{F_0}(f)\big), \quad (7)$$

where $ш_{F_0}(f)$ is a Dirac comb with fundamental frequency $F_0 = 1/T_0$, and $\widetilde{n_g}$ and $\widetilde{n'_g}$ are the Fourier transforms of n_g and n'_g respectively. It must be emphasized that $\widetilde{n_g}$ and $\widetilde{n'_g}$ depend only on α_m in these equations. Therefore the spectral effect of parameters E, T_0 and O_q can be studied independently of any particular model: they will have the same effect whichever the GFM.

According to equation (7), $\widetilde{U'_g}$ is derived from $\widetilde{n'_g}$ by the following operations: a frequency scaling by O_qT_0, an amplitude scaling by EO_qT_0 and a sampling of this resulting continuous spectrum, which is the harmonic envelope, with a sampling rate of F_0 to obtain the harmonics. Note that the periodic repetition by T_0 results in a sampling of the spectrum by F_0 but introduces also an amplitude change of F_0 because the Fourier transform of $ш_{T_0}(t)$ is $F_0 ш_{F_0}(f)$.

The influence of the different parameters on $\widetilde{U'_g}(f;P)$ can be deduced from equations (6) and (7):

- E has the effect of an overall gain.
- T_0 allows the whole spectrum to stretch or shrink, the harmonic amplitudes and phases being unchanged.
- O_q has the same type of effect as T_0 but it stretches or shrinks only the spectral envelope, without changing the harmonic frequencies.
- the effect of α_m depends on the specific generic model used. This will be discussed below.

The effect of the return phase quotient is considered in paragraph 3.4. The spectra of the generic form of glottal flow models are derived in Appendix A2, where analytic expressions for the Laplace transform for each of the four models considered here are elaborated.

3.3. Glottal formant

The first definition of the glottal formant is due to Fant [13]: "The glottal pulse frequency F_g is defined as the inverse of twice the duration of the rising branch": $F_g = 1/2T_p$. This definition has the advantage of simplicity but, paradoxically, is a time domain definition which represents a spectral feature of only one part of the glottal pulse and does not define the glottal formant amplitude. We adopt a sligthly different definition, starting from a frequency domain point of view: in this paper, the glottal formant is defined as the maximum of the glottal flow derivative spectrum. It is to be noticed that both definitions coincide if they are applied on the LF model with abrupt closure but with a non-truncated open phase towards the negative times. It has been shown that the truncation of the damped sinusoid in the LF model shifts the frequency of the maximum slightly upward [21]. This is consistent with Fant's observation that "the sinusoidal residue of the glottal pulse (...) appears in the spectrum as a "baseband" formant at or somewhat above F_g" [13].

Applying the same procedure as in paragraph 3.1 to the preceding equations, it can be shown that the GFM spectrum in its generic form actually behaves as a low-pass filter frequency response. The glottal-formant frequency and amplitude can then be deduced.

When the frequency tends towards 0, the GFM spectrum tends towards a constant which corresponds to the GFM integral I. Since GFM waveforms are always positive, this constant is non-null. The spectrum at frequency 0 is then given by:

$$\widetilde{U_g}(0;P) = I \quad (8)$$

When the frequency tends to infinity, the asymptotic properties of the spectra are linked to the discontinuity in the derivative of the GFM at the glottal closing instant. It can be shown that if there are no other discontinuities in the GFM derivative than the E gap at GCI (in particular if there is no gap at opening time), then the GFM spectrum behaves at infinity as a $1/f^2$ (i.e. -12 dB/oct) slope with amplitude E

$$|\widetilde{U_g}(f;P)| \overset{f\to+\infty}{\sim} \frac{E}{(2\pi f)^2} \quad (9)$$

Then, the asymptote (i.e. the behaviour of the spectra for high frequencies) does not depend on T_0, O_q nor even on α_m. Therefore it is independent of the particular model used. E is the only parameter that controls the medium and high frequency spectral behaviour of a GFM with abrupt closure.

Consider also the GFM derivative. The GFM derivative spectrum is the product of the GFM spectrum by $j2\pi f$. This product transforms the slopes of the asymptotic lines as follows. The spectrum is equivalent to $I2\pi f$ when the frequency tends towards 0, and therefore the slope is +6 dB/oct. The spectrum is equivalent to $E/2\pi f$ when the frequency tends towards infinity, and therefore the slope is −6 dB/oct.

Thus the spectrum of the GFM derivative behaves like a bandpass filter frequency response. Figure 8 shows the spectral magnitude envelope (amplitude gain) of such a filter. The spectral peak of this filter can be characterized by its peak frequency (corresponding to the cut-off frequency of the GFM spectrum) F_g and its peak amplitude A_g defined by the crossing point between the two asymptotic lines

$$F_g = \frac{1}{2\pi}\sqrt{\frac{E}{I}}, \tag{10}$$

$$A_g = \sqrt{EI}. \tag{11}$$

Equation (10) shows that the spectral peak frequency is determined by the amplitude of the derivative discontinuity, over the GFM integral. In other words, this frequency depends on the speed of closure of the vocal folds over the total glottal flow. Consider now how F_g and A_g are related to the time-domain generic parameters.

For that, the spectrum of the generic model derivative $\widetilde{n'_g}$ must be considered. This spectrum shows the same type of asymptotic behaviour (−6 dB/oct, +6 dB/oct), with a crossing point defined by a frequency $f_g(\alpha_m)$ and an amplitude $a_g(\alpha_m)$ that can be explicitly calculated from the expression of n'_g. All the equations have been reported in Appendix A2 for the sake of clarity of the main text.

Using equations (10), (11) and (3), it can be deduced that

$$F_g = \frac{1}{O_qT_0}\frac{1}{2\pi\sqrt{i_n(\alpha_m)}} = \frac{f_g(\alpha_m)}{O_qT_0} = \frac{f_g(\alpha_m)F_0}{O_q}, \tag{12}$$

$$A_g = EO_qT_0\sqrt{i_n(\alpha_m)} = EO_qT_0a_g(\alpha_m). \tag{13}$$

However what we are interested in is the glottal formant, which is defined as the maximum of the GFM derivative spectrum (see Figure 8). It must be pointed out that, as for the second-order linear filter, the actual spectrum maximum is not exactly at the frequency position of the asymptote crossing point. The true maximum is not found at frequency F_g but at a slightly higher frequency depending on the specific equation of the generic model and thus on the asymmetry coefficient. Using equation (7), the frequency F_{max} and amplitude A_{max} of the glottal formant can be written as a function of the frequency $f_{max}(\alpha_m)$ and the amplitude $a_{max}(\alpha_m)$ of the maximum of $\widetilde{n'_g}$ as

$$\begin{aligned} F_{max} &= \arg\max_f \widetilde{U'_g}(f;P) = \frac{1}{O_qT_0}\arg\max_f \widetilde{n'_g}(f;\alpha_m) \\ &= \frac{1}{O_qT_0}f_{max}(\alpha_m), \end{aligned} \tag{14}$$

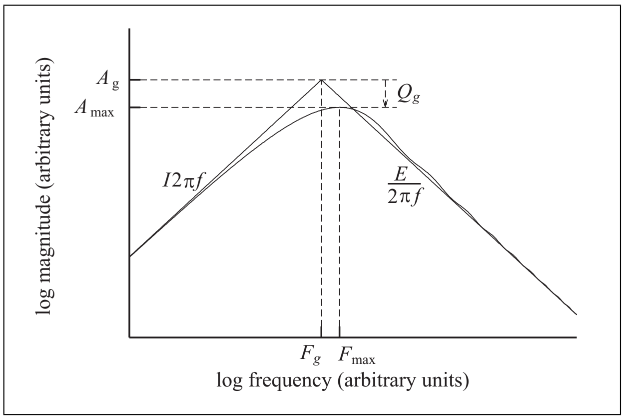

Figure 8. Glottal formant. In the abrupt closure case, the GFM derivative spectrum can be stylized by 2 asymptotes. The asymptote in $f = 0$ and in $f = \infty$ are controlled by the total flow I and the maximum excitation E resp. Their crossing point is (F_g, A_g). This gives raise to the "glottal formant" the frequency of which F_{max} is slightly higher than F_g. The difference Q_g between A_g and A_{max} is a way to characterize the glottal formant bandwidth.

$$\begin{aligned} A_{max} &= \max_f \widetilde{U'_g}(f;P) = EO_qT_0\max_f \widetilde{n'_g}(f;\alpha_m) \\ &= EO_qT_0a_{max}(\alpha_m). \end{aligned} \tag{15}$$

These equations show that F_{max} and A_{max} are influenced by O_q or T_0 in exactly the same way as F_g and A_g. But contrary to the asymptote crossing point, for which analytical expressions are available, the maximum of the derivative spectrum can not be analytically and explicitly related to α_m, at least for the 4 studied models. Although explicit expressions for $f_{max}(\alpha_m)$ and $a_{max}(\alpha_m)$ have not been found, they would still be of interest because f_{max} and a_{max} are functions of only one parameter which can be easily computed by a numerical algorithm.

The influence of the asymmetry coefficient α_m can be studied using the preceding equations. The difference between the stylized spectral envelope defined by the two asymptotic lines and the actual spectral envelope is mainly due to the bandwidth of the glottal spectral peak. This bandwidth can be modelled by a quality coefficient Q_g which measures the difference in dB between A_{max} and A_g, in a way that is analogous to second-order linear filters. Substituting these values by their expression given by equations (13) and (15) gives

$$Q_g = \frac{A_{max}}{A_g} = \frac{a_{max}(\alpha_m)}{a_g(\alpha_m)} = q_g(\alpha_m), \tag{16}$$

showing that this coefficient is independent of E, T_0 and O_q, and that it is only related to the asymmetry coefficient α_m. Figure 14 represents the spectral changes corresponding to several values of α_m for the LF GFM , and Figure 9 shows the function $q_g(\alpha_m)$ for the 4 studied GFMs.

Returning to Fant's original spectral model (1960), it is interesting to note that it also showed a glottal formant at a frequency that can be controlled by the position of the poles s_{r1} and s_{r2} but with a constant bandwidth. This can be compared with the more recent GFMs which control the glottal formant bandwidth through the asymmetry coefficient.

Finally, one must remember that the actual source spectrum is a harmonic spectrum, whose spectral envelope is defined by the spectrum of one GFM pulse. This fact is illustrated in Figure 10.

3.4. Spectral tilt

In the preceding paragraphs, only the abrupt closure case has been considered. Here we consider GFMs with smooth closure ($Q_a \neq 0$). The spectral effect of the GCI time-domain smoothing is mainly an attenuation of the high frequencies as noted by Fant et al. (1985, p. 8): "Even a very small departure from abrupt termination causes a significant spectrum roll-off in addition to the standard -12 dB/oct glottal flow spectrum". This large spectral change for small parametric variation can be explained by the change in nature of the wave (differentiable versus non-differentiable). The term spectral tilt is very often used to designate this spectrum roll-off.

Historically, in the early spectral model of [1], two fixed poles s_{r3} and s_{r4} were attenuating high frequencies of the voice source. Then, in the Rosenberg time-domain models [3], there was no longer any high-frequency attenuation. But Fant [2], Klatt [4], Veldhuis [5] and others reintroduced the spectral tilt parameter either in the time domain or in the frequency domain. This is because this parameter is of the utmost importance for voice quality. According to [22], the spectral tilt component is one of the main spectral cues for prosodic stress perception.

For GFMs with smooth closure, it has been seen in paragraph 2.1 that two methods were used: either the return phase method or the low-pass filter method. In the low-pass filter method, the smooth closure GFM spectrum is obtained from the abrupt closure GFM spectrum by multiplying it by the transfer function of a first or second order filter. For a first order filter, this is given by

$$LPF(s) = \frac{1}{1 + \frac{s}{2\pi F_c}} \tag{17}$$

where F_c is the cut-off frequency of the filter. The impulse response of this filter is a decreasing exponential with time constant T_c given by $T_c = 1/(2\pi F_c)$. As has been stated before, this time constant is very similar to the return phase parameter T_a and can be used as an approximation of it. But this filter is more often specified by the attenuation TL (in dB) at a given medium frequency (for instance $3000Hz$ for the KLGLOTT88 model). Then, from the equations above, TL can be related to T_a by: $T_a = \sqrt{10^{TL/10} - 1}/(2\pi 3000)$. The main advantage of the low-pass filter method is that the low frequency spectral features, especially the glottal formant, are kept constant or only slightly changed in a predictable way.

In the return phase method, according to [2], the main spectral effect of the return phase is an additional low-pass filtering with a cut-off frequency given by:

$$F_c \simeq F_a = \frac{1}{2\pi T_a} \tag{18}$$

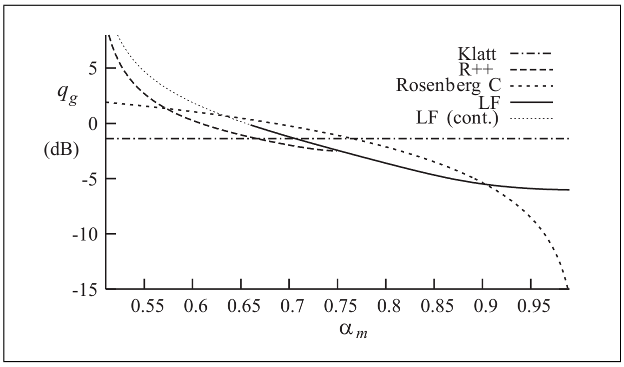

Figure 9. Quality coefficient q_g in function of α_m for the 4 GFMs. Note that q_g depends only on the parameter α_m. For the LF-model it is plotted with a dashed line for values of α_m lower than 0.65.

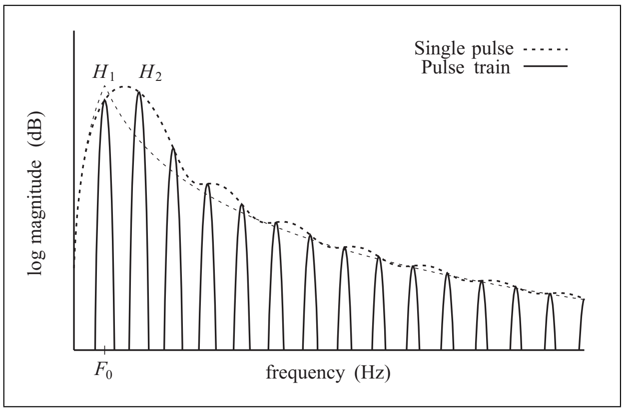

Figure 10. Spectrum of a pulse train for the LF model.

This approximation can be compared with the above approximation $T_a \simeq T_c$ for the low-pass filter method. For the LF model, a more precise approximation can be obtained by using the equation [23]: $F_c = F_a + a/(2\pi) + R_g F_0 cot(\pi(1 + R_k))$. But the main drawback of the return phase method with a constrained baseline is that it also changes the low frequencies. This is due to the time-domain modification of the open phase implied by the additional return phase.

Thus, for any abrupt closure GFM, both smoothing methods can be used and have the same type of effect in time and frequency. The usual parameters (T_a or TL) can be related to the return phase quotient by the above equations. Figure 13 shows the effect of the spectral tilt component in the GFM spectrum. The right side of glottal formant is modified by an additional -6 dB/oct, after a frequency point defined by F_c. The spectral tilt factor can be -6 dB/oct or more depending on the corresponding order of the spectral tilt filter.

3.5. Phase spectrum

More and more attention is paid nowadays to the phase characteristics of the source and its usage for inverse filtering [24] or speech coding. The source phase has been shown to contain information on the source parameters,

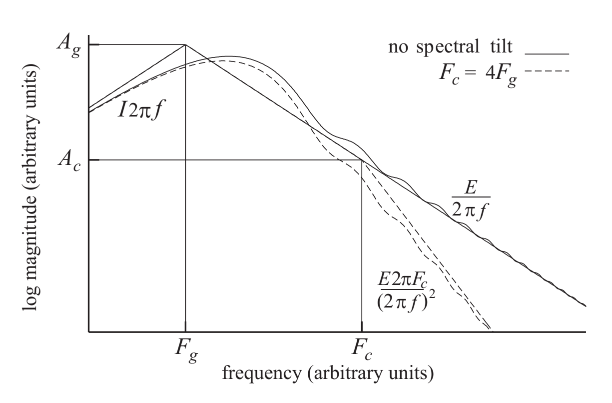

Figure 11. Spectrum and asymptotes of the GFM derivative spectrum with (dashed) and without (plain) spectral tilt. An additional −6 dB/oct slope is added above frequency F_c.

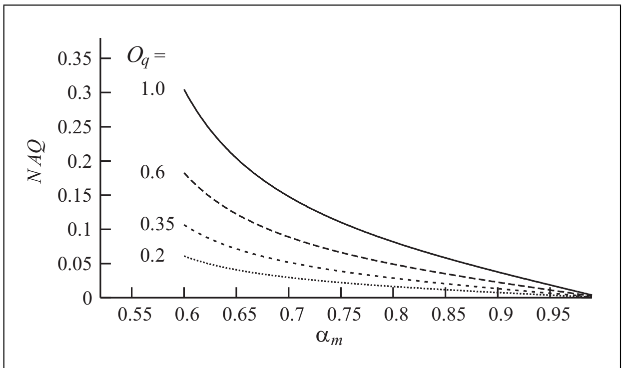

Figure 12. Parameter NAQ in function of α_m for different values of O_q, in the case of abrupt closure and for the LF model. $NAQ = A_v/ET_0 = a_v(\alpha_m)O_q$.

and especially those related to voice quality. Some insight is then needed into the structure of GFM phase spectra.

There is some evidence that part of the phase spectrum exhibits some anticausal behaviour [25, 23]. This anticausal behaviour can be related in the time domain to the right skewness of the open phase. Returning to the second order filter analogy, the open phase looks like the impulse response of an anticausal second order filter. In the frequency domain, the anticausality cannot be viewed on the magnitude spectrum because both versions (causal and anticausal) have the same magnitude spectrum. But the GFM phase spectrum exhibits increasing values from low to middle frequencies in the region corresponding to the glottal formant. This can also be observed on the group delay where the delay becomes negative around the glottal formant frequency (see Figure 7). However the return phase part of the GFM exhibits a decreasing phase spectrum and a positive group delay around the cut-off frequency.

Finally, the open phase of the GFM is anticausal while the return phase is causal, resulting in a “mixed phase” model [26, 21]. This anticausality property can be used for inverse filtering purposes to estimate the open phase from speech signals or to estimate the frequency of the glottal formant [27].

3.6. Glottal spectral parameters

In summary, the spectral envelope of glottal flow models can be considered as the gain of a lowpass filter. The spectral envelope of the GFM derivative can then be considered as the gain of a bandpass filter. Linear stylization of the spectrum in a log-log representation is presented in Figure 11. The spectrum of any GFM derivative can be stylized by 3 linear segments with +6 dB/oct, −6 dB/oct and −12 dB/oct (or sometimes −18 dB/oct) slopes, respectively. The 2 breakpoints correspond to the glottal formant frequency and the spectral tilt cut-off frequency. Their frequency and amplitude are F_g, A_g, F_c, A_c. The amplitude A_c can be deduced from F_g, A_g and F_c.

At this point, the idea of describing a GFM by spectral parameters should be considered. A set of such parameters could then be: the fundamental frequency F_0, the amplitude A_g and frequency F_g of the first breakpoint, the quality coefficient Q_g of the glottal formant, and the frequency F_c of the spectral tilt breakpoint. The advantage of this set is that there are exact formulae to relate those parameters to the time-domain generic parameters. The drawback is that it is not easy to estimate or even to observe the breakpoints on the source spectrum. Another set would then be obtained by replacing A_g and F_g by the amplitude A_{max} and frequency F_{max} of the glottal formant. These parameters are easier to estimate but are not analytically related to the time-domain parameters. However, as has been stated, numerical algorithms can be used to obtain them from the equations. Finally, the parameter E could be used in place of A_g or A_{max} since it has a clear spectral role (it defines the position of the second line in the stylization).

In the remainder of the paper, the results obtained in this section are used for studying the spectral correlates of the time-domain parameters.

4. Spectral correlates of time-domain parameters

In this section, the role played by time-domain parameters and their spectral consequences are discussed in detail, with emphasis on the low-frequency part of the spectrum related to the glottal formant. Further, the relationship between open quotient and the first two-harmonics amplitude difference is revisited here, as this amplitude is often used in the literature to estimate the open quotient.

4.1. Spectral correlates of amplitude and return phase quotient

The amplitude E of the negative peak of the glottal flow derivative, also called maximum excitation, is the GFM amplitude parameter in the time domain. The amplitude of voicing A_v is an alternative amplitude parameter (see Figure 13 and sound example 6). Both parameters behave

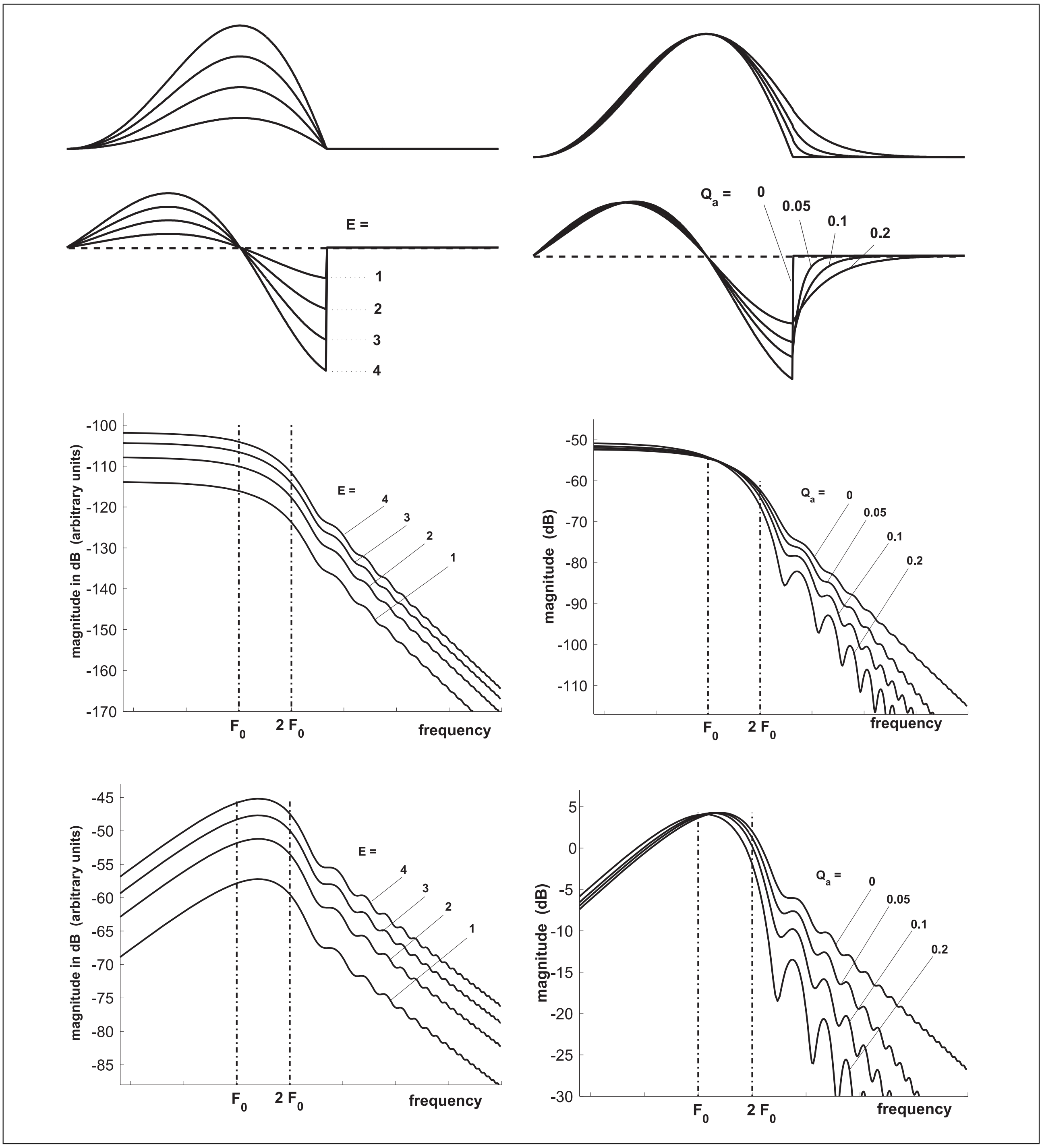

Figure 13. Correlates of E / A_v (left) and Q_a (right) on the GFM and the GFM derivative spectra (LF model). E or A_v plays the role of a gain. Q_a controls the spectral tilt and modifies slightly the low-frequency region (A_v being fixed).

as a gain factor in the glottal flow or glottal flow derivative spectra. All the equations can be written using E or A_v. As shown above, E controls the -6 dB/oct part of the spectrum, which resides between the position of the glottal formant and the spectral-tilt cut-off frequency point. This encompasses the mid frequency harmonics and also, depending on the other parameters, the low and/or high frequency harmonics. This could explain the high correlation between E and the SPL found in many studies such as [28, 29, 30]. For that matter, when the other parameters are varied, keeping E or A_v constant is not equivalent. When considering the effect of an open quotient variation, it is important to specify which amplitude parameter is kept constant. For instance, the just noticeable differences (JND) of O_q variations measured when E is kept constant are approximately twice the ones measured when A_v is kept constant [31]. This can be heard by comparing sound examples 9 and 11.

Spectral correlates of the return phase quotient are shown in Figure 13 and sound example 7. Its main effect

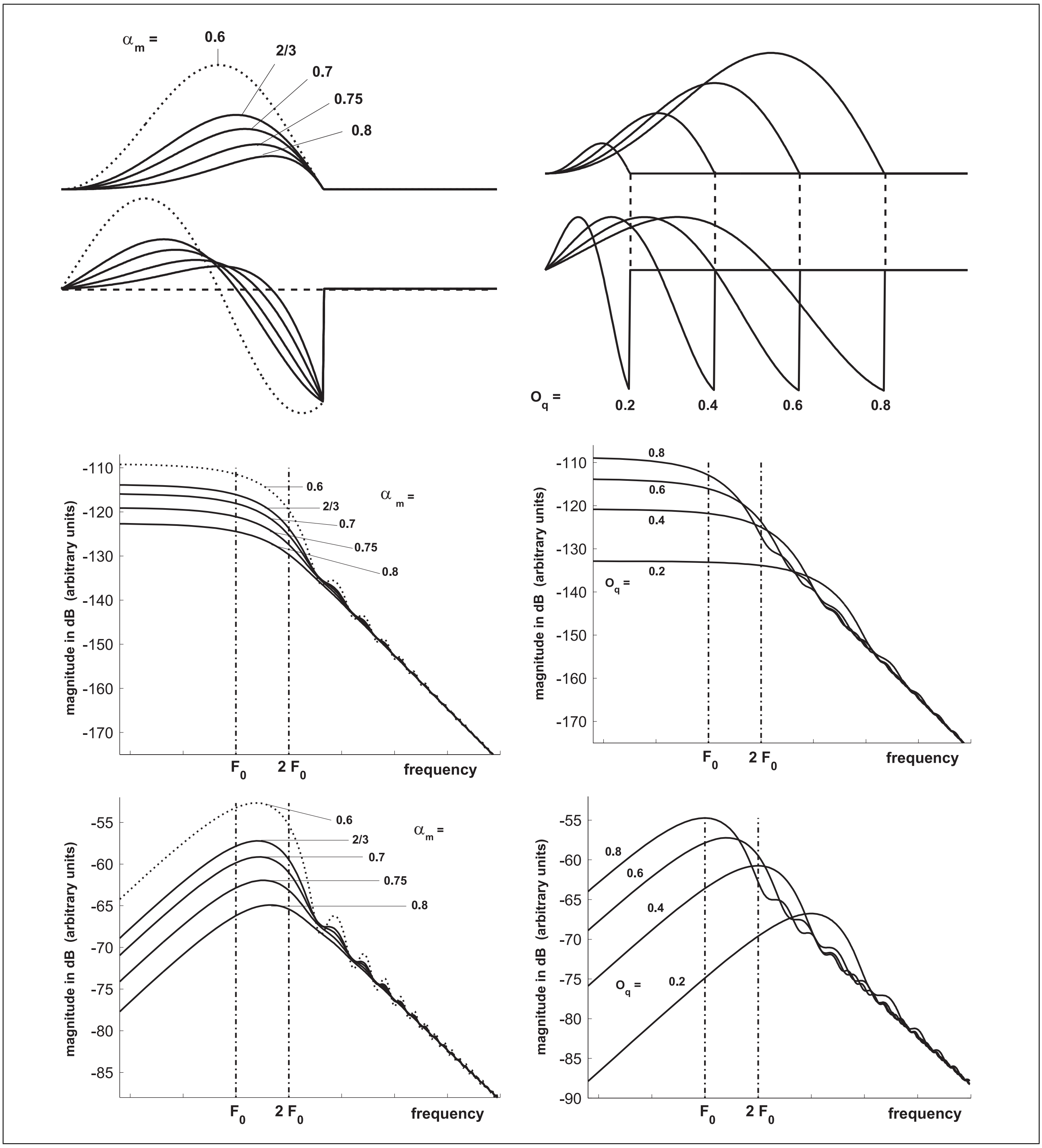

Figure 14. Correlates of α_m (left) and O_q (right) on the GFM and the GFM derivative spectra, E being fixed (LF model). O_q and α_m mainly affect the low-frequency region of the spectrum, More precisely α_m controls the glottal formant bandwidth and O_q its frequency. Note that the variation of the spectrum maximum magnitude (A_{max}) is a consequence of the concomitant variation in A_v.

is to introduce an additional spectral tilt above the cut-off frequency. This spectral turning point can easily be computed (see section 3.4). The return phase quotient can also influence slightly the centre frequency and bandwidth of the glottal formant, but this seems to be only a second order effect. In the return-phase implementation of the spectral tilt, the return phase takes place between O_qT_0 and T_0. This implies that O_q must be strictly lower than 1 and that the return phase quotient must be bounded by 1. Therefore the cut-off frequency is lower-bounded by the value corresponding to $Q_a = 1$. This lower bound depends on the particular GFM. When Q_a is high (near 1) the cut-off frequency is low and the spectral tilt is high, leading to the high frequency attenuation observed in soft voices. In contrast, when Q_a is low, the cut-off frequency is high and the spectral tilt is low, this case corresponding to loud voices. There is no upper bound for the cut-off frequency since Q_a can be as low as required and can even be null, in which

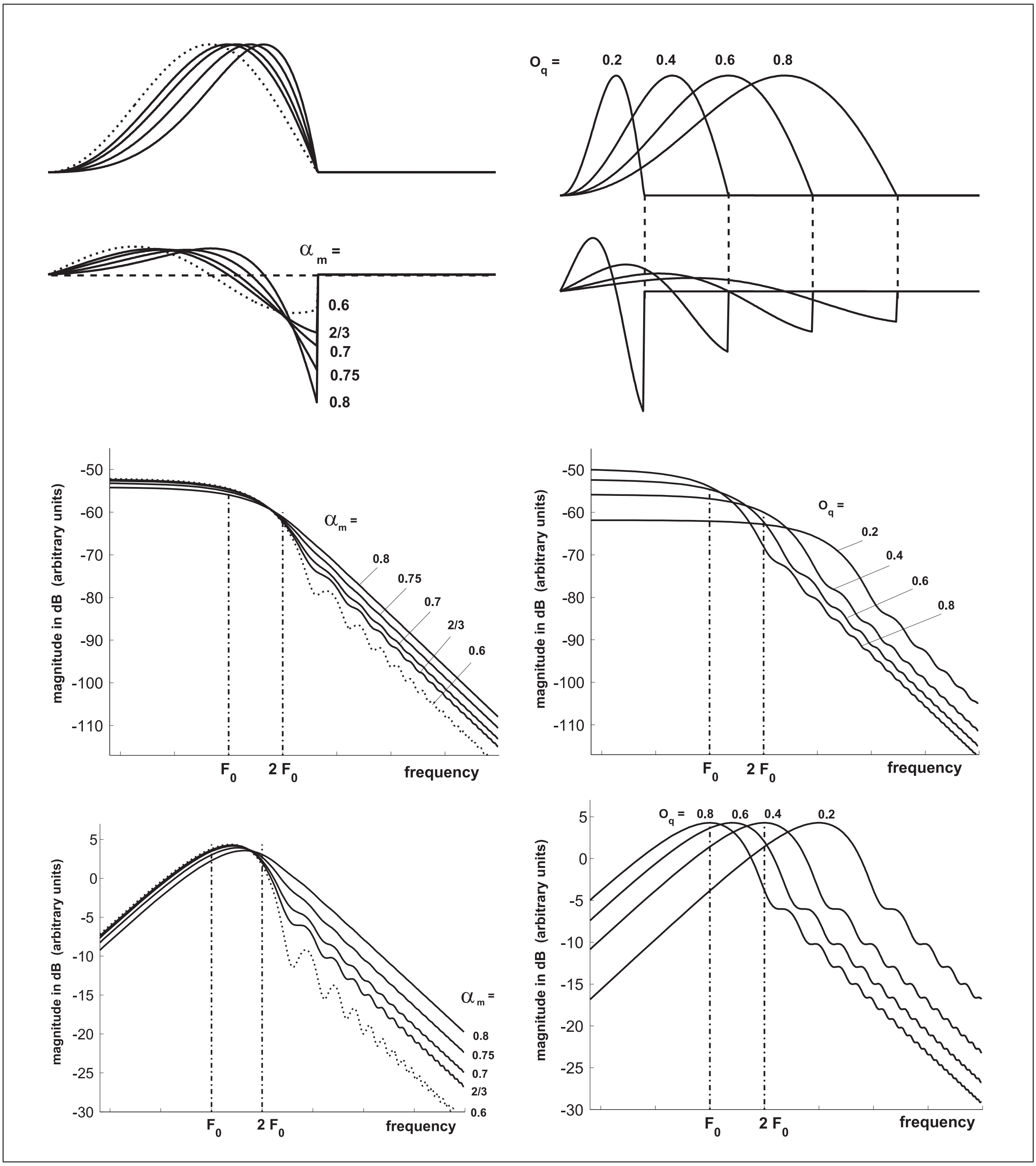

Figure 15. Correlates of O_q and α_m on the GFM and the GFM derivative spectra, A_v being fixed (LF model). When E is not fixed, a variation of O_q or α_m can change the medium and high frequency regions.

case the closure is abrupt, the cut-off tends towards infinity and there is no additional attenuation in the spectrum.

4.2. Effect of O_q and α_m on the glottal formant

The influence of α_m and O_q on the glottal formant is represented in Figure 14, where LF GFM spectra are plotted for 4 values of α_m and 4 values of O_q, in the abrupt closure case ($Q_a = 0$) and while E is kept constant. This figure shows that the main effect is indeed in the low frequency region and that the mid and high frequencies are not much modified by α_m and O_q variations. This point will now be discussed in more detail.

First, as can be seen in equations (12) and (14), the asymptote crossing frequency F_g and the glottal formant maximum frequency F_{max} are proportional to F_0. It is thus more interesting to study its position relative to the harmonics rather than to give absolute values.

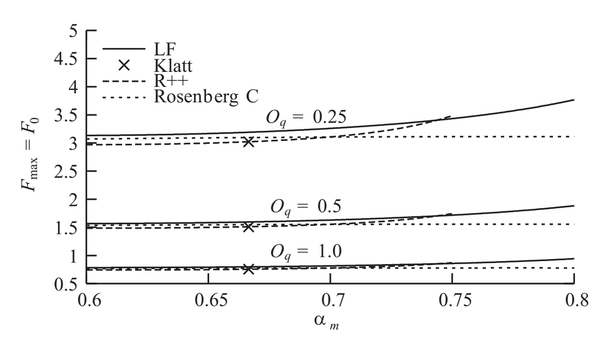

Figure 16. Glottal formant frequency: F_{max}/F_0 in function of α_m for different values of O_q and different models. Note that it is almost independent of α_m (with small variations along models) but is inversely proportional to O_q (see equation 14).

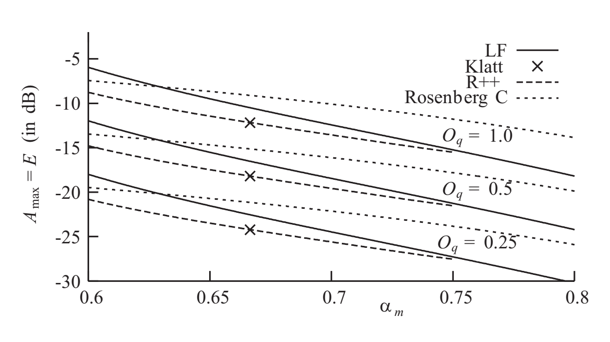

Figure 17. Glottal formant magnitude: A_{max}/E in function of α_m for different values of O_q and different models. Note that it decreases with α_m (with variations along models) and is proportional to O_q (see equation 15).

Thus, Figure 16 shows the variation of F_{max}/F_0 as a function of α_m and O_q, for standard values ($\alpha_m = 0.6, \ldots, 0.8$, $O_q = 0.25, \ldots, 1.0$) and for the four GFM. As can be seen, this ratio is only slightly affected by a change of α_m, but is inversely proportional to O_q, as shown in equation 14.

The values of F_{max}/F_0 range from 0.74 to 3.8 and this extent shows very small variations from one model to another: for O_q values ranging from 0.25 to 1.0 and α_m values from 0.6 to 0.8 (but following the limitations of each model), this ratio shows a range of $[0.76, 3.0]$ for Klatt, $[0.74, 3.5]$ for R++, $[0.77, 3.1]$ for C, and $[0.78, 3.8]$ for LF. These bounds indicate that the GFM derivative spectrum reaches its maximum roughly between the first and the fourth harmonics. The lower bound is a theoretical bound due to the fact that O_q cannot be greater than 1. There is no theoretical upper bound and the given upper bound is fixed by the lowest considered value of O_q and, to a lesser extent, by the highest considered value of α_m. Notice that Fant's original definition of the glottal formant frequency follows the same behaviour: it is proportional to F_0, lower bounded (by $F_0/2$) and not upper bounded.

In Figure 17 the relative magnitude A_{max}/E of the spectral maximum for each model is plotted with respect to α_m and O_q for the same standard values as mentioned above and with a reference value of $T_0 = 1$. As can be seen in this Figure, the maximum magnitude is proportional to O_q as given by equation (15), and then increases by 6 dB each time O_q is multiplied by 2, given that E is constant. We observe in this case a variation of 12 dB for an O_q range of $[0.25, 1.0]$. The maximum magnitude is also greatly influenced by α_m: a variation of -12.2 dB is observed for the LF model when α_m moves from 0.6 to 0.8. This variation is even greater for extreme values of α_m, i.e. $\alpha_m > 0.8$. This influence is model-dependent and smaller for the R++ model (-6.7 dB for $\alpha_m = 0.6$ to 0.75) and the C model (-6.4 dB for $\alpha_m = 0.6$ to 0.8).

Reconsidering the second-order filter analogy for the glottal formant, the influence of α_m should be compared to that of a quality factor which controls the bandwidth, or equivalently the difference between the maximum and the asymptote crossing point amplitudes. This is clearly shown in Figure 14: α_m seems to play the role of the glottal-formant bandwidth while O_q controls its position.

Following [4], when the open quotient decreases, the voice quality changes from a lax to a pressed voice. From the observations presented above, we can say that in this case the glottal formant moves from a low towards a mid frequency, while attenuating mainly the first-harmonics amplitude, the higher harmonics being hardly changed (sound example 9). The voice quality modification due to the asymmetry coefficient (or equivalently the speed quotient) is rarely described in the literature. From the observations presented above, we can say that it mainly affects the relative amplitude of the first harmonics, having a similar effect to the open quotient on the tense-lax axis (sound example 8).

In that direction, Alku *et al.* [32] proposed the parameter NAQ defined by the T_0-normalized ratio between A_v and E as a way to parametrize the glottal closing phase. Equation 3 shows that for GFMs this parameter can be directly related to O_q and α_m: $NAQ = a_v(\alpha_m)O_q$. Since $a_v(\alpha_m)$ is a decreasing function of α_m (see Appendix A2 for formula), NAQ increases with O_q and decreases with α_m (Figure 12). This could explain why it is considered to capture the relative degree of tenseness/laxness.

The major difference between α_m and O_q spectral effects is that α_m can be used to boost the harmonics in the vicinity of the glottal formant. For instance, in Figure 14, the first harmonic is boosted by 12 dB when α_m goes from 0.8 to 0.6. Conversely, for α_m values greater than 0.8, the effect of an α_m variation is very similar to that of an O_q variation. When $\alpha_m > 0.9$, it can even be simulated by a variation of O_q.

Figure 15 shows the spectral influence of O_q and α_m while A_v is kept constant (instead of E). As can be seen, this influence is a combination of the low frequency effect described above and the overall magnitude change due to the variations of E, which can be seen on the high frequency asymptote (sound examples 10 and 11).

For the low-pass filter implementation of the spectral tilt, its influence on the glottal-formant maximum frequency is negligible and its influence on the maximum magnitude is to add an extra 0 to −3 dB amplitude if we consider that the cut-off frequency is always greater than F_{max}. But for the return phase case it seems that, in some particular cases, the influence is greater and cannot be considered as negligible. This happens especially when the return phase quotient is high (near 1) and the open quotient low: in this case the large area under the return phase in the GFM-derivative waveform is compensated for by a modification of the waveform before the GCI, and this leads to some changes in the spectrum around the glottal-formant frequency. However this particular case is hardly seen in real speech or singing as it corresponds to a very soft voice (Q_a is high) which is also pressed (O_q is low).

4.3. First two harmonic amplitudes vs. O_q and α_m

In many studies [33, 7, 34, 34, 35], the spectral amplitude difference between the first two harmonics is assumed to be strongly correlated to the open quotient. It is often used as a spectral measure of the open quotient. In this part, we shall discuss this assumption, on the basis of the theoretical framework presented above. Let $H_1 - H_2$ denote the difference between the first two harmonic amplitudes (in dB), as measured on the glottal flow derivative spectrum. These spectral differences can be measured either in the inverse-filtered voiced signal or in the voiced signal itself by using a formant-based correction according to [7, 34].

In the case of abrupt closure, Fant [15] found a very good correlation between $H_1 - H_2$ and O_q, using the LF model. This correlation is almost perfectly linear according to the equation:

$$H_1 - H_2 = -6 + 0.27 * \mathrm{e}^{5.5 O_q} \tag{19}$$

Then, according to this equation, $H_1 - H_2$ depends on the open quotient alone. However, when one computes harmonic amplitudes according to equation (7), one obtains:

$$\begin{aligned} H_1 - H_2 &= 20 \log_{10} \left| \widetilde{U'_g}(F_0; P) \right| \\ &\quad - 20 \log_{10} \left| \widetilde{U'_g}(2F_0; P) \right| \\ &= 20 \log_{10} \left| \frac{\widetilde{n'_g}(O_q; \alpha_m)}{\widetilde{n'_g}(2O_q; \alpha_m)} \right| . \end{aligned} \tag{20}$$

This equation shows that $H_1 - H_2$ does not depend on the fundamental frequency or the amplitude parameter but on both the open quotient and the asymmetry coefficient.

While studying glottal characteristics in case of male [36] and female speakers [7, 34], Hanson estimated O_q with measurements of $H_1 - H_2$. Her work is based on the KLGLOTT88 model. In this model, the asymmetry coefficient is of constant value $\alpha_m = 2/3$. Therefore, there is a direct relationship between O_q and $H_1 - H_2$, as illustrated in Figure 18. A unique value of $H_1 - H_2$ corresponds to a given value of O_q, at least considering only values of $O_q \leq 0.74$. Without this hypothesis, for $H_1 - H_2$ between 7.94 and 9.82, two values of O_q give the same harmonic differences.

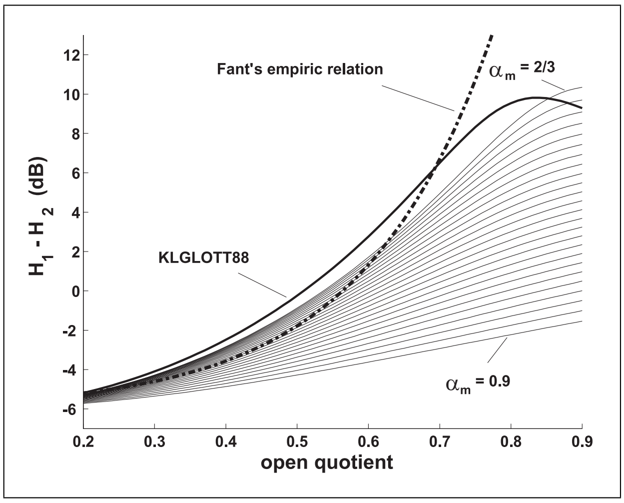

Figure 18. $H_1 - H_2$ in function of O_q for the Klatt model and for several values of α_m for the LF model. The Fant's empirical relation between $H_1 - H_2$ and O_q (see equation 19) has been superimposed.

Contrary to the KLGLOTT88 model, the asymmetry coefficient can be varied in the LF, R++ or Rosenberg C models. Thus, harmonic differences depend on both the open quotient and the asymmetry coefficient for those models. In Figure 18, $H_1 - H_2$ is plotted as a function of the open quotient for different values of asymmetry coefficient, ranging from $\alpha_m = 2/3$ to $\alpha_m = 0.9$ with steps of 0.01. A given value of $H_1 - H_2$ does not correspond to an unique value of open quotient, but to an open quotient interval as a function of the α_m. For instance, Hanson [34] found an average value of 3.4 dB for $H_1 - H_2$ (mean value over 22 female speakers, vowel /æ/). For this average value, the possible open quotient values range between 0.66 and 1.0 and the possible asymmetry coefficient values between 2/3 and 0.81. Several couples of parameters, e.g. ($O_q = 0.66$, $\alpha_m = 2/3$), ($O_q = 0.80$, $\alpha_m = 0.77$) or ($O_q = 1.0$, $\alpha_m = 0.81$) would give a same value for $H_1 - H_2$ according to the LF model.

These figures show that for one value of $H_1 - H_2$, 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 the case of smooth closure (higher spectral tilt). It seems that the parameter Q_a has indeed an effect on the measure $H_1 - H_2$, but that this effect is much less important than the effect of the open quotient and asymmetry coefficient, and can thus be neglected.

5. Conclusion

The aim of this paper is to study the spectrum of glottal flow models (GFM). For this, a unified framework has been established to easily derive analytical formulae for the spectra derived from four widely used GFMs (LF, Klatt, R++ and Rosenberg C).

First, it has been shown that the GFMs can be described by a common set of 5 time-domain parameters: the fundamental period, the maximum excitation, the open quotient, the asymmetry coefficient and the return phase coefficient. All the GFMs are equivalent with regard to the first 3 parameters but the last 2 parameters are model specific, even if they produce very similar effects among the models.

Also, the spectra from all GFMs have been shown to be equivalent to that of a low-pass filter. More precisely, the spectra from the GFM derivatives exhibits a spectral peak at low frequency that is called the glottal formant and an additional attenuation at medium or high frequency corresponding to the spectral tilt. The structure of the glottal formant has been shown to be that of a second order anticausal resonant filter with an overall stylized shape of +6 dB/oct then −6 dB/oct. Analytical expressions for the glottal formant frequency and amplitude have been given as a function of time-domain parameters. They show that the glottal formant frequency is mainly controlled by the open quotient while its amplitude (or equivalently its bandwidth) is mainly controlled by the asymmetry coefficient. For standard values of time-domain parameters, the glottal formant roughly takes place between the first and the fourth harmonic. Moreover, the −6 dB/oct part of the spectrum is uniquely controlled by the maximum excitation, which can explain the high correlation between this parameter and the SPL shown by [30]. The spectral tilt part of the spectrum behaves like a first (or second) low-pass filter which results in a −12 dB/oct (or −18 dB/oct) slope in the GFM derivative spectrum. The cut-off frequency of this filter is mainly related to the return phase coefficient.

As a direct application, the relationship between the difference of the two first harmonic amplitudes and the open quotient has been explored and this difference has been shown to be also dependent on the asymmetry coefficient.

Many other applications could benefit from these theoretical results. First, since the glottal formant and spectral tilt parameters have been shown to be equivalent to the time-domain parameters, they can be used to control a spectrally defined GFM [21]. Another application is the spectral estimation of glottal flow parameters [37, 38] that would also take into account the phase information (which has been shown to be important [39, 25]). This would help for voice quality analysis [34, 22]. Finally, spectral domain modification of the voice source [40] could also be developed. However, the extent to which modifications based only on linear filtering would be sufficient, remains to be tested in more detail.

Appendix

A1. Generic model for four glottal flow models

In this appendix, four classical GFMs are reviewed. For each model, the relationship between the original model parameters and the generic parameters is given.

A1.1. KLGLOTT88 model

This model derived from the Rosenberg B [3] model has been used in the Klatt synthesizer [4], and in several studies (e.g. [7, 34]).

The glottal waveform U_g [4] is characterized by four parameters: the fundamental frequency $F_0 = 1/T_0$, the amplitude of voicing AV, the open quotient O_q, and the attenuation of a spectral tilt filter TL. When the parameter TL is set to 0 dB, the KLGLOTT88 model is identical to the Rosenberg B model. In this case, the equation of the model and its derivative are:

$$U_{g_{TL=0}}(t) = \begin{cases} at^2 - bt^3 & 0 \le t \le O_q T_0, \\ 0 & O_q T_0 \le t \le T_0, \end{cases}$$

with $a = \frac{27}{4}\frac{AV}{O_q^2 T_0}$, $b = \frac{27}{4}\frac{AV}{O_q^3 T_0^2}$,

$$U'_{g_{TL=0}}(t) = \begin{cases} 2at - 3bt^2 & 0 \le t \le O_q T_0, \\ 0 & O_q T_0 \le t \le T_0. \end{cases}$$

The generic parameters are: T_0, O_q, $\alpha_m = 2/3$ and $E = -U'_{g_{TL=0}}(O_q T_0) = 27AV/4O_q$. The generic GFM is then obtained by taking: $T_0 = 1$, $O_q = 1$ and $A = 4/27$.

When $TL \neq 0$, $U_{g_{TL=0}}(t)$ is filtered by a first or second-order low-pass filter, such as the attenuation at 3000 Hz equals TL dB. A side effect of the low-pass filter is to shift in time the maximum of the model. However the GCI itself is not shifted by a first order filter.

A1.2. R++ model

Veldhuis [5] recently revisited the Rosenberg B model, and proposed the so-called R++ model. This model is improved along to axes an asymmetry coefficient and a return phase parameter.

The model proposed by [5] is defined using 5 parameters:

- K: amplitude coefficient.
- T_0: fundamental period.
- T_e: minimum of the glottal flow derivative waveform (excitation instant).
- T_p: maximum of the glottal flow waveform.
- T_a: time constant for the return phase.

The glottal flow derivative model contains a 3rd order polynomial part (between time 0 and T_e), followed by an exponential return phase (between T_e and T_0):

$$U'_g(t) = \begin{cases} 4Kt(T_p - t)(T_x - t) & 0 \le t \le T_e, \\ U'_g(T_e)\frac{e^{-(t-T_e)/T_a} - e^{-(T_0-T_e)/T_a}}{1-e^{-(T_0-T_e)/T_a}} & T_e < t \le T_0. \end{cases}$$

In this equation, parameter T_x is computed as:

$$T_x = T_e\left(1 - \frac{\frac{1}{2}T_e^2 - T_e T_p}{2T_e^2 - 3T_e T_p + 6T_a(T_e - T_p)D(T_0, T_e, T_a)}\right),$$

with

$$D(T_0, T_e, T_a) = 1 - \frac{(T_0 - T_e)/T_a}{e^{(T_0-T_e)/T_a} - 1}$$

The meaning of parameter T_x is as follows. The GFM must fulfil some conditions: 1. the waveform is null for $t = 0$; 2. the derivative of the GFM is null for $t = 0$; 3. the maximum of GFM is for $t = T_p$; 4. the derivative is continuous for $t = T_e$; 5. the GFM is null for $t = T_0$. Conditions 2., 3. and 4. are satisfied by the definition of $U_g'(t)$. But conditions 1. and 5. are involving that the integral of the model derivative between 0 and T_0 is null. This condition gives an equation which in turn defines T_x.

The R++ GFM is obtained by integrating its derivative:

$$U_g(t) = \begin{cases} Kt^2\left(t^2 - \frac{4}{3}t\left(T_p + T_x\right) + 2T_pT_x\right) \\ \qquad\qquad 0 \le t \le T_e, \\ U_g(T_e) + T_aU_g'(T_e) \\ \quad \cdot \frac{1 - e^{-(t-T_e)/T_a} - ((t-T_e)/T_a)e^{-(T_0-T_e)/T_a}}{1 - e^{-(T_0-T_e)/T_a}} \\ \qquad\qquad T_e \le t \le T_0. \end{cases}$$

The maximum excitation is given by: $U_g'(T_e) = 4KT_e(T_p - T_e)(T_x - T_e)$. If $T_a = 0$, then T_x reduces to: $T_x = T_e(3T_e - 4T_p)/2(2T_e - 3T_p)$. The parameter E is then equal to $E = -U_g'(T_e) = 2KT_e^2(T_e - T_p)(2T_p - T_e)/(2T_e - 3T_p)$. The other generic parameters are T_0, $O_q = T_e/T_0$, $\alpha_m = T_p/T_e$. Then the generic model and its derivative are obtained taking $T_0 = 1$, $T_e = 1$, $T_p = \alpha_m$, and $K = (2 - 3\alpha_m)/[2(2\alpha_m - 1)(1 - \alpha_m)]$.

A1.3. Rosenberg-C

The Rosenberg C [3] model is a trigonometric model, defined using 4 parameters:

- A: amplitude.
- T_0: fundamental period.
- T_p: maximum of the glottal flow waveform.
- T_n: time interval between maximum of the glottal flow waveform and the GCI (thus $T_p + T_n$ is the GCI).

The model is made of two sinusoidal parts:

$$U_g(t) = \begin{cases} \frac{A}{2}(1 - cos(\pi\frac{t}{T_p})) & 0 \le t \le T_p \\ Acos(\frac{\pi}{2}\frac{t-T_p}{T_n}) & T_p \le t \le T_p + T_n \\ 0 & T_p + T_n \le t \le T_0 \end{cases}$$

the glottal flow derivative is:

$$U_g'(t) = \begin{cases} \frac{\pi A}{2T_p}(sin(\pi\frac{t}{T_p})) & 0 \le t \le T_p \\ -\frac{\pi A}{2T_n}sin(\frac{\pi}{2}\frac{t-T_p}{T_n}) & T_p \le t \le T_p + T_n \\ 0 & T_p + T_n \le t \le T_0 \end{cases}$$

The generic parameters are obtained from the original parameters as: $O_q = (T_p + T_n)/T_0$, $\alpha_m = T_p/(T_p + T_n)$ and $E = \pi A/2T_n$. The generic GFM is then obtained by taking: $T_0 = 1$, $T_p = \alpha_m$, $T_n = 1 - \alpha_m$ and $A = 2/\pi\,(1 - \alpha_m)$.

A1.4. LF model

The LF model [2] represents the glottal flow derivative. Its 5 parameters are defined in the time domain:

- E_e: amplitude at the minimum of the glottal flow derivative (maximum of excitation).
- T_0: fundamental period.
- T_e: instant of maximum excitation.
- T_p: instant of the maximum of the glottal flow.
- T_a: time constant of the return phase.

This model is made of a sinusoidal part modulated by a rising exponential (between 0 and T_e), followed by an decreasing exponential return phase (between T_e and T_0):

$$U_g'(t) = \begin{cases} -E_e e^{a(t-T_e)}\frac{sin(\pi t/T_p)}{sin(\pi T_e/T_p)} & 0 \le t \le T_e \\ \frac{-E_e}{\varepsilon T_a}(e^{-\varepsilon(t-T_e)} - e^{-\varepsilon(T_0-T_e)}) & T_e \le t \le T_0 \end{cases}$$

In this equation, the parameter ε is defined by an implicit equation:

$$\varepsilon T_a = 1 - e^{-\varepsilon(T_0-T_e)}$$

This equation can be solved for computing ε, provided that T_a, T_e and T_0 are known. The parameter a is defined by an implicit equation:

$$\frac{1}{a^2 + (\frac{\pi}{T_p})^2}\left(e^{-aT_e}\frac{\pi/T_p}{sin(\pi T_e/T_p)} + a - \frac{\pi}{T_p}cotg(\pi T_e/T_p)\right) = \frac{T_0 - T_e}{e^{\varepsilon(T_0-T_e)} - 1} - \frac{1}{\varepsilon}$$

This equation can be solved for computing a, provided that T_p, T_e, T_0, and ε are known.

The LF model is a model of the glottal flow derivative. As the glottal flow must be null for $t = 0$ and $t = T_0$, the integral of the glottal flow derivative between 0 and T_0 must be null. This necessary condition explains the implicit equation for a. Except for the constant factor $-E_e$, the left member of the equation is the integral of $U_g'(t)$ between 0 and T_e, and the right member of the equation is the opposite of the integral of $U_g'(t)$ between T_e and T_0.

The implicit equation for ε is derived from the continuity of the waveform at T_e. It is also possible, by combining the two equations, to give an expression of ε as a function of a and the model parameters.

The open quotient is defined by:

$$O_q = \frac{T_e}{T_0}$$

As this definition does not take into account the return phase duration, Fant proposed the following modification:

$$O_q = \frac{T_e + T_a}{T_0}$$

The glottal flow is obtained by integrating its derivative $U_g'(t)$:

$$U_g(t) = \begin{cases} \frac{-E_e e^{-aT_e}}{sin(\pi T_e/T_p)}\frac{1}{a^2 + (\frac{\pi}{T_p})^2}\Big(\frac{\pi}{T_p} + ae^{at}sin(\pi t/T_p) \\ \qquad -\frac{\pi}{T_p}e^{at}cos(\pi t/T_p)\Big) \\ \qquad\qquad 0 \le t \le T_e, \\ -E_e(\frac{1}{\varepsilon T_a} - 1)(T_0 - t + \frac{1}{\varepsilon}(1 - e^{\varepsilon(T_0-t)})) \\ \qquad\qquad T_e \le t \le T_0. \end{cases}$$

When $T_a = 0$, the generic parameters are given by: T_0, $O_q = T_e/T_0$, $\alpha_m = T_p/T_e$ and $E = E_e$. The generic GFM is then obtained taking $T_0 = 1$, $T_e = 1$, $T_p = \alpha_m$, and $E_e = 1$.

A2. Spectra of four glottal flow models

Let $\mathcal{L}(n_g')(s)$ be the Laplace transform and $\widetilde{n_g'}(f)$ the Fourier transform of the generic GFM derivative $n_g'(t)$. Both transforms always exist because all the considered waveforms are of finite duration. Moreover, the Fourier transform is obtained as the Laplace transform taken on the imaginary axis: $\widetilde{n_g'}(f) = \mathcal{L}(n_g')(j2\pi f)$.

In the following are given, for the four studied GFMs, the analytic expressions of the corresponding generic GFMs, their Laplace transform, their generic amplitude of voicing, their generic total flow, and the coordinates of their asymptote crossing point.

A2.1. KLGLOTT88 model

The generic KLGLOTT88 model and its derivative are given by

$$n_g(t) = (t^2 - t^3) \tag{A1}$$
$$n_g'(t) = (2t - 3t^2) \tag{A2}$$

Then the shape factors are:

$$\alpha_m = \frac{2}{3}, \qquad a_v = \frac{4}{27}, \qquad i_n = \frac{1}{12}.$$

It is noticeable that the shape parameter α_m is fixed for the KLGLOTT88 model. This model has only 4 degrees of freedom.

The Laplace transform of the KLGLOTT88 model is given by

$$\mathcal{L}(n_g')(s) = \frac{1}{s}\left(e^{-s} + \frac{2(1+2e^{-s})}{s} - \frac{6(1-e^{-s})}{s^2}\right) \tag{A3}$$

and the glottal spectral peak of the generic model is:

$$f_g = \frac{\sqrt{3}}{\pi}, \qquad a_g = \frac{1}{2\sqrt{3}}.$$

It must be pointed out that for this model the glottal spectral peak depends only on the open quotient. This is because the shape parameter is fixed in the KLGLOTT88 model.

The spectral tilt component is a first or second order low-pass filter, whose frequency is chosen in order to obtain a given attenuation at 3 kHz.

A2.2. R++ model

The generic GFM of the R++ model is given by

$$n_g(t;\alpha_m) = \frac{2-3\alpha_m}{2(1-\alpha_m)(2\alpha_m-1)}t^2(t-1) \cdot \left(t - \frac{\alpha_m(3-4\alpha_m)}{2-3\alpha_m}\right), \tag{A4}$$

$$n_g'(t;\alpha_m) = \frac{2-3\alpha_m}{(1-\alpha_m)(2\alpha_m-1)}2t(t-\alpha_m) \cdot \left(t - \frac{3-4\alpha_m}{2(2-3\alpha_m)}\right). \tag{A5}$$

α_m is a free parameter, and the shape factors are given by:

$$a_v(\alpha_m) = \frac{\alpha_m^3(1-\alpha_m)}{2(2\alpha_m-1)},$$
$$i_n(\alpha_m) = \frac{-3+12\alpha_m-10\alpha_m^2}{60(1-\alpha_m)(2\alpha_m-1)}.$$

It must be noted that the range of α_m must be restricted to $[0.5, 0.75]$, otherwise the GFM may be negative.

The Laplace transform of the R++ model is given by

$$\begin{aligned}\mathcal{L}(n_g')(s;\alpha_m) = {} & \frac{1}{(1-\alpha_m)(2\alpha_m-1)}\frac{1}{s} \\ & \cdot \Big[e^{-s}(1-\alpha_m)(2\alpha_m-1) + \frac{1}{s}\big(\alpha_m(3-4\alpha_m) \\ & - e^{-s}(8\alpha_m^2 - 15\alpha_m + 6)\big) \\ & - \frac{6}{s^2}\big(1 - 2\alpha_m^2 + e^{-s}(2\alpha_m^2 - 6\alpha_m + 3)\big) \\ & + \frac{12}{s^3}(1-e^{-s})(2-3\alpha_m)\Big]\end{aligned}$$

and the glottal spectral peak of the generic model is:

$$f_g(\alpha_m) = \frac{1}{2\pi}\sqrt{\frac{60(1-\alpha_m)(2\alpha_m-1)}{-3+12\alpha_m-10\alpha_m^2}},$$
$$a_g(\alpha_m) = \sqrt{\frac{-3+12\alpha_m-10\alpha_m^2}{60(1-\alpha_m)(2\alpha_m-1)}}.$$

The spectral tilt component is computed using the return phase method.

A2.3. Rosenberg-C model

The generic GFM of the Rosenberg C model is given by

$$n_g(t;\alpha_m) = \left| \begin{array}{ll} \frac{1-\alpha_m}{\pi}(1 - cos(\pi\frac{t}{\alpha_m})) & 0 \le t \le \alpha_m, \\ \frac{2(1-\alpha_m)}{\pi} sin(\frac{\pi}{2(1-\alpha_m)}(1-t)) & \alpha_m \le t \le 1, \end{array} \right. \tag{A6}$$

$$n_g'(t;\alpha_m) = \left| \begin{array}{ll} (\frac{1}{\alpha_m} - 1) sin(\pi\frac{t}{\alpha_m}) & 0 \le t \le \alpha_m, \\ -cos(\frac{\pi}{2(1-\alpha_m)}(1-t)) & \alpha_m \le t \le 1. \end{array} \right. \tag{A7}$$

α_m is a free parameter, and the shape factors are given by:

$$a_v(\alpha_m) = \frac{2}{\pi}(1-\alpha_m),$$
$$i_n(\alpha_m) = (\frac{2}{\pi})^2(1 - \alpha_m(1-\frac{\pi}{4}))(1-\alpha_m).$$

The Laplace transform of the Rosenberg C model is given by

$$\mathcal{L}(n_g')(s;\alpha_m) = \frac{2}{\pi}(1-\alpha_m)\left(\frac{1}{1+(\frac{\alpha_m}{\pi}s)^2}\frac{1+e^{-\alpha_m s}}{2} + \frac{\frac{2}{\pi}(1-\alpha_m)se^{-s} - e^{-\alpha_m s}}{1+\left(\frac{2}{\pi}(1-\alpha_m)s\right)^2}\right), \tag{A8}$$

and the glottal spectral peak of the generic model is:

$$f_g(\alpha_m) = \frac{1}{4\sqrt{(1-\alpha_m(1-\pi/4))(1-\alpha_m)}},$$
$$a_g(\alpha_m) = \frac{2}{\pi}\sqrt{(1-\alpha_m(1-\pi/4))(1-\alpha_m)}.$$

A2.4. LF model

The generic GFM of the Liljencrants-Fant (LF) model and its derivative are

$$n_g(t;\alpha_m) = \frac{\pi/\alpha_m}{-\mathrm{e}^{a_n} sin(\pi/\alpha_m)(a_n^2+(\pi/\alpha_m)^2)} \cdot \left(1+\mathrm{e}^{a_n t}\left(a_n \frac{\alpha_m}{\pi} sin(\pi t/\alpha_m) - cos(\pi t/\alpha_m)\right)\right), \quad \text{(A9)}$$

$$n'_g(t;\alpha_m) = -\mathrm{e}^{a_n(t-1)} \frac{sin(\pi t/\alpha_m)}{sin(\pi/\alpha_m)}, \quad \text{(A10)}$$

where a_n must satisfy the implicit equation $n_g(1) = 0$, i.e.:

$$1+\mathrm{e}^{a_n}\left(a_n \frac{\alpha_m}{\pi} sin\left(\frac{\pi}{\alpha_m}\right) - cos\left(\frac{\pi}{\alpha_m}\right)\right) = 0. \quad \text{(A11)}$$

a_n is similar to the parameter a of the complete GFM and the relationship between them is: $a_n = aO_qT_0$.

α_m is a free parameter, and the shape factors are given by:

$$a_v(\alpha_m) = \frac{\frac{\pi}{\alpha_m}(1+\mathrm{e}^{a_n\alpha_m})}{-\mathrm{e}^{a_n} sin(\pi/\alpha_m)(a_n^2+(\pi/\alpha_m)^2)},$$

$$i_n(\alpha_m) = \frac{1-\frac{\pi/\alpha_m}{\mathrm{e}^{a_n} sin(\pi/\alpha_m)}}{a_n^2+(\pi/\alpha_m)^2}.$$

It must be noted that when α_m is lower than 0.65, the negative peak of the GFM derivative is no longer at time T_e and its amplitude can be very different from $-E$. For instance, with $\alpha_m = 0.55$, the negative peak value is $-1.8E$, and with $\alpha_m = 0.51$, it increases to $-8.1E$. This can be important for source parameter estimation procedures based on time-domain estimation of the glottal-flow derivative maximum. However the proposed framework can handle the full α_m range ($]0.5, 1.0[$) as can be seen in figure 14 where the asymptote remains $E/(2\pi f)$ whatever the value of α_m.

The Laplace transform of the LF model is given by

$$\mathcal{L}(n'_g)(s;\alpha_m) = \frac{s\mathrm{e}^{-s} - \frac{\pi/\alpha_m}{sin(\pi/\alpha_m)}\mathrm{e}^{-a_n}(1-\mathrm{e}^{-s})}{(a_n-s)^2+(\pi/\alpha_m)^2}, \quad \text{(A12)}$$

and the glottal spectral peak of the generic model is:

$$f_g(\alpha_m) = \frac{1}{2\pi}\sqrt{\frac{a_n^2+(\pi/\alpha_m)^2}{1-\frac{\pi/\alpha_m}{\mathrm{e}^{a_n} sin(\pi/\alpha_m)}}},$$

$$a_g(\alpha_m) = \sqrt{\frac{1-\frac{\pi/\alpha_m}{\mathrm{e}^{a_n} sin(\pi/\alpha_m)}}{a_n^2+(\pi/\alpha_m)^2}}.$$

The spectral tilt component is computed using the return phase method.

List of sounds

All the sounds are computed using an LF GFM derivative filtered by a vowel /a/ filter, except when notified.

Sound 1: LF GFM, then GFM derivative, then GFM derivative through a vowel filter. $F_0 = 100$, $O_q = 0.6$, $\alpha_m = \frac{2}{3}$, $Q_a = 0$.

Sound 2: same as sound 1, but with a series of F_0 and E values taken from a natural vowel.

Sound 3, 4, 5: Comparison between models: effect of α_m variation for the C, LF and R++ GFMs (respectively sounds 3, 4 and 5). α_m moves from 0.6 to 0.8 and back for the C and LF GFMs, and from 0.55 to 0.75 and back for the R++ GFM. $E = 1000$, $F_0 = 130$, $O_q = 0.6$, $Q_a = 0$.

Sound 6: Effect of E variation (see Figure 13). E moves from 1000 to 4000 and back, first by steps of 1000 and then continuously. $F_0 = 130$, $O_q = 0.6$, $\alpha_m = \frac{2}{3}$, $Q_a = 0$.

Sound 7: Effect of Q_a variation (see Figure 13). Q_a takes the values 0, 0.05, 0.1, 0.2 and back, and then moves from 0 to 0.2 and back continuously. $AV = 1$, $F_0 = 130$, $O_q = 0.6$, $\alpha_m = \frac{2}{3}$.

Sound 8: Effect of α_m variation with fixed E (see Figure 14). α_m takes the values $\frac{2}{3}$, 0.7, 0.75, 0.8 and back, and then moves from $\frac{2}{3}$ to 0.8 and back continuously. $E = 1000$, $F_0 = 130$, $O_q = 0.6$, $Q_a = 0$.

Sound 9: Effect of O_q variation with fixed E (see Figure 14). O_q takes the values 0.8, 0.6, 0.4, 0.2 and back, and then moves from 0.8 to 0.2 and back continuously. $E = 1000$, $F_0 = 130$, $\alpha_m = \frac{2}{3}$, $Q_a = 0$.

Sound 10: Effect of α_m variation with fixed A_V (see Figure 15). α_m takes the values $\frac{2}{3}$, 0.7, 0.75, 0.8 and back, and then moves from $\frac{2}{3}$ to 0.8 and back continuously. $A_V = 1$, $F_0 = 130$, $O_q = 0.6$, $Q_a = 0$.

Sound 11: Effect of O_q variation with fixed A_V (see Figure 15). O_q takes the values 0.8, 0.6, 0.4, 0.2 and back, and then moves from 0.8 to 0.2 and back continuously. $A_V = 1$, $F_0 = 130$, $O_q = 0.6$, $\alpha_m = \frac{2}{3}$, $Q_a = 0$.

References

[1] G. Fant: Acoustic theory of speech production. Mouton, The Hague, 1960.

[2] G. Fant, J. Liljencrants, Q. Lin: A four-parameter model of glottal flow. STL-QPSR **4** (1985) 1–13.

[3] A. E. Rosenberg: Effect of glottal pulse shape on the quality of natural vowels. J. Acous. Soc. Am. **49** (1971) 583–590.

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

[5] R. Veldhuis: A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation. J. Acous. Soc. Am. **103** (1998) 566–571.

[6] H. Fujisaki, M. Ljungqvist: Proposal and evaluation of models for the glottal source waveform. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 1986.

[7] H. M. Hanson: Glottal characteristics of female speakers. Dissertation. Harvard University, 1995.

[8] D. G. Childers, C. K. Lee: Vocal quality factors: Analysis, synthesis, and perception. J. Acous. Soc. Am. **90** (1991) 2394–2410.

[9] P. Alku, H. Strik, E. Vilkman: Parabolic spectral parameter - a new method for quantification of the glottal flow. Speech Communication **22** (1997) 67–79.

[10] L. C. Oliveira: Estimation of source parameters by frequency analysis. Proc. Eurospeech'93, 1993, 99–102.

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

[12] B. Doval, C. d'Alessandro, B. Diard: Spectral methods for voice source parameter estimation. Proc. Europseech'97, Rhodes, Greece, Sept. 1997, 533–536.

[13] G. Fant: Glottal source and excitation analysis. STL-QPSR **1** (1979) 85–107.

[14] T. V. Ananthapadmanabha: Acoustic analysis of voice source dynamics. STL-QPSR **2-3** (1984) 1–24.

[15] G. Fant: The LF-model revisited. transformations and frequency domain analysis. STL-QPSR **2-3** (1995) 119–156.

[16] N. Henrich, C. d'Alessandro, M. Castellengo, B. Doval: On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. J. Acous. Soc. Am. **115** (2004) 1321–1332.

[17] I. R. Titze: Theory of glottal airflow and source-filter interaction in speaking and singing. Acta Acustica united with Acustica **90** (2004) 641–648.

[18] N. Henrich, C. d'Alessandro, B. Doval: Spectral correlates of voice open quotient and glottal flow asymmetry : theory, limits and experimental data. Eurospeech 2001, Aalborg, Denmark, Sept. 2001.

[19] N. Henrich: Etude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électroglottographiques, perception (study of the glottal source in speech and singing: modeling and estimation, acoustic and electroglottographic measurements, perception). Dissertation. Université Paris 6, France, 2001.

[20] J. D. Markel, A. H. Gray: Linear prediction of speech. Springer-Verlag, Berlin, 1976.

[21] B. Doval, C. d'Alessandro: The voice source as a causal / anticausal linear filter. proc. Voqual'03, Voice Quality: Functions, analysis and synthesis, ISCA workshop, Geneva, Switzerland, Aug. 2003, 15–20.

[22] A. Sluijter, V. J. van Heuven, J. J. A. Pacilly: Spectral balance as a cue in the perception of linguistic stress. J. Acous. Soc. Am. **101** (1997) 503–513.

[23] B. Doval, C. d'Alessandro: Spectral correlates of glottal waveform models : an analytic study. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Munich, Germany, Apr. 1997, 446–452.

[24] P. Alku, M. Airas, T. Bäckström, H. Pulakka: Using group delay function to assess glottal flows estimated by inverse filtering. IEE, Electronics Letters **41** (Apr. 2005).

[25] W. Gardner, B. D. Dao: Non-causal all-pole modelling of voiced speech. IEEE Trans. On Speech and Audio Proc. **5** (1997) 1–10.

[26] B. Bozkurt: Zeros of z-transform(zzt) representation and chirp group delay processing for analysis of source and filter characteristics of speech signals. Dissertation. Université Polytechnique de Mons, Belgium and LIMSI-CNRS, France, 2005.

[27] B. Bozkurt, B. Doval, C. d'Alessandro, T. Dutoit: A method for glottal formant frequency estimation. 8th Intern. Conf. on Spoken Language Processing (ICSLP), Jeju Island, Korea, Okt. 2004.

[28] G. Fant: Preliminaries to analysis of the human voice source. STL-QPSR **4** (1982) 1–27.

[29] E. B. Holmberg, R. E. Hillman, J. S. Perkell: Glottal air flow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice. J. Acous. Soc. Am. **84** (1988) 511–529.

[30] J. Gauffin, J. Sundberg: Spectral correlates of glottal voice source waveform characteristics. J. Speech Hear. Res. **32** (1989) 556–565.

[31] N. Henrich, G. Sundin, D. Ambroise, C. d'Alessandro, M. Castellengo, B. Doval: Just noticeable differences of open quotient and asymmetry coefficient in singing voice. J. Voice (2003).

[32] P. Alku, T. Bäckström, E. Vilkman: Normalized amplitude quotient for parametrization of the glottal flow. J. Acous. Soc. Am. **112** (Aug. 2002) 701–710.

[33] E. B. Holmberg, R. E. Hillman, J. S. Perkell, P. C. Guiod, S. L. Goldman: Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. J. Speech Hear. Res. **38** (1995) 1212–1223.

[34] H. M. Hanson: Glottal characteristics of female speakers : Acoustic correlates. J. Acous. Soc. Am. **101** (1997) 466–481.

[35] J. Sundberg, M. Andersson, C. Hultqvist: Effects of subglottal pressure variation on professional baritone singers' voice sources. J. Acous. Soc. Am. **105** (1999) 1965–1971.

[36] H. M. Hanson, E. S. Chuang: Glottal characteristics of male speakers : Acoustic correlates and comparison with female data. J. Acous. Soc. Am. **106** (1999) 1064–1077.

[37] L. C. Oliveira: Text-to-speech synthesis with dynamic control of source parameters. – In: Progress in Speech Synthesis. J. van Santen, R. Sproat, J. Olive, J. Hirschberg (eds.). Springer-Verlag, 1996, 27–39.

[38] B. Bozkurt, B. Doval, C. d'Alessandro, T. Dutoit: Zeros of z-transform representation with application to source-filter separation in speech. IEEE Signal processing letters **12** (Apr. 2005) 344–347.

[39] B. Bozkurt, B. Doval, C. d'Alessandro, T. Dutoit: Appropriate windowing for group delay analysis and roots of z-transform of speech signals. 12th European Signal Processing Conference (EUSIPCO), Vienna, Austria, Sep. 2004.

[40] C. d'Alessandro, B. Doval: Voice quality modification using periodic-aperiodic decomposition and spectral processing of the voice source signal. Proc. 3rd Intern. Workshop on Speech Synthesis, Jenolan Caves, Australia, Nov. 1998, 277–282.