Stylization of glottal-flow spectra produced by a mechanical vocal-fold model

Denisse Sciamarella, Christophe D’Alessandro

LIMSI-CNRS
Université de Paris-Sud 91403 ORSAY (France)
sciamarella@limsi.fr

Abstract

A method is proposed to extract glottal-flow spectra from numerical simulations of vocal-fold behavior with a two-mass model including dynamic flow separation. The numerical spectrum, whose general form complies with that of signal glottal-flow models, allows stylization with three linear segments. The slope of the first segment remains relatively constant when source control parameters are varied, whereas the slope of the last segment (i.e. the spectral tilt) is highly sensitive to the vibrating vocal-fold mass, tension and stiffness. The phase of mobility of the flow separation position within the glottal cycle may introduce, if long enough, a dip in the glottal-flow spectrum.

1. Introduction

The voice source is the origin of all the features of speech related to voice quality, vocal effort, and prosodic variations. The possibility of characterizing glottal activity through time-domain physiological analyses (such as electroglottography, ultra-rapid photography or electromyography) strongly encouraged time-domain modeling of the glottal pulse or waveform, also called flow glottogram ([1, 2, 3]). However, when it comes to describing vocal quality, the glottal-flow spectrum is believed to be more suitable than the glottogram itself (as has been thoroughly noted in most applications involving speech analysis and synthesis). According to [4] for instance, one of the main spectral parameters for synthesizing voices with different qualities is the rate of decay of the voice source spectrum (known as spectral slope or spectral tilt). A number of studies focus on this issue : [5, 6, 7, 8].

A unified framework for studying the time and frequency domain properties of glottal flow models has been proposed in [9]. In this work, the authors show that the glottal-flow spectra can be stylized by 3 straight lines in a log-magnitude/log-frequency representation characterized by five frequency-domain parameters : fundamental frequency, spectral peak amplitude and frequency, quality factor of the spectral peak and spectral tilt cut-off frequency.

The aim of this paper is to inspect, by means of numerical simulations, in which manner the adjustment of the mechanical properties of the vocal folds and subglottal pressure influence the glottal-flow spectrum (see [10] for an exhaustive study on the glottal-flow waveform). The numerical method we present consists in computing the Fourier transform for the glottal pulse derivative obtained by simulating vocal fold behavior with a recent two-mass model. The procedure is completed by measuring the three slopes stylizing the spectrum. The results of the numerical experiments give some insight into the dynamic role of the vocal source behavior on voice quality.

2. The method

The production model used in the simulations is the so called Niels Lous two-mass model, whose main features are the inclusion of dynamic flow separation within the glottal channel and the assumption of a symmetrical glottal structure [11] (i.e. the punctual masses m₁ and m₂ in Figure 1 are assigned the same value m). Such a model is chosen for its conceptual simplicity and its well known rich behavior.

The main speaker’s control parameters in our model are subglottal pressure P₁, vocal fold tension k, the tension k, coupling the lower and upper parts of the vocal folds - also called vocal-fold stiffness, the vocal fold vibrating mass m and length L₀. Typical values for these parameters are : m ≈ 0.1 g, k ≈ 40 N/m, kₗ ≈ 25 N/m, L₀ ≈ 1.4 cm and P₁ ≈ 8 cmH₂O. These typical values are certainly not static in speech. They make part of the active control parameters that the speaker can change or adjust. For instance, L₀ can be stretched in 3 or 4 mm during phonation.

We will hereafter focus on the repercussion of each of these parameters on the shape of the glottal-flow spectrum. With this purpose, control parameters will be set to adopt a number of values p within a physiologically meaningful range in order to compute the glottal flow Uₚ(t) (as well as its derivative, yₚ(t), yₚ(t), etc.) for each p value. An algorithmic procedure will extract a glottal-flow pulse sample from each time series Uₚ(t) (after excluding transients) and analyse Uₚ(t) to compute the time domain and frequency domain parameters.

The magnitude of the discrete Fourier transform \(|\mathcal{F}(Uₚ(t))|\) of the glottal-flow derivative is numerically computed and as follows :
where $\mathcal{F}$ indicates Fourier transform, $N$ is the length of the time series, frequency $f = \frac{s}{2N\Delta t}$ with $1 < s < N + 1$ and $\Delta t = T_0/N = 1/S_r$ (with $S_r$ : sampling rate). The result of this calculation is next submitted to an algorithmic procedure in order to extract the parameters characterizing its general form.

For a brief definition of these parameters, consider Figure 2, which shows the spectrum of a trigonometric signal glottal-flow model [12] with abrupt glottal closure. The spectral magnitude (in dB) is defined as $10 \log(|\mathcal{F}(U_g^p)(s)|)$. The log-log spectrum has a peak at $(F_m, A_m)$ and an asymptotic behavior in $6\text{dB/oct}$ and $-6\text{dB/oct}$ when frequency tends respectively to zero and infinity. This type of asymptotic behavior can be characterized by the frequency of the crossing point between the two asymptotic lines, which gives birth to an approximate spectral peak $F_g$. If vocal fold closure is smooth, the glottal-flow model becomes differentiable at closure. Every departure from abrupt closure causes a spectrum roll-off in addition to the standard glottal-flow spectrum [2], introducing a spectral tilt cut-off frequency $F_t$. The spectrum can be hence stylized by three linear segments, with breakpoints at $(F_g, A_g)$ and $(F_c, A_c)$ ([9]). Since the magnitude $A_c$ can be derived from $F_g$, $A_g$ and $F_c$, the set of five parameters defining the spectrum according to [9] is $F_0, A_g, F_g, Q_g = |A_g - A_m|, F_c$.

The algorithmic routine used to compute these parameters from $|\mathcal{F}(U_g^p)(s)|$ consists in calculating first the position of the actual spectral peak $F_m$. Three linear regressions are then performed for: (i) the first half of the data lying before $F_m$ (line $L_1$), (ii) the last half of the data lying after $F_m$ (line $L_2$), (iii) the first half of the data lying after $F_m$ (line $L_3$). If $L_2 \cap L_3 \neq \emptyset$, then $F_c$ is computed from $L_2 \cap L_3$ and $(F_g, A_g) = L_1 \cap L_3$. Otherwise, $F_c$ will be set to zero and the stylization will be done exclusively with $L_1$ and $L_2$, so that $(F_g, A_g) = L_1 \cap L_2$. It is clear that for glottal-flow signals with an abrupt closure, $L_2 \rightarrow L_3$. Finally, an output file stores the computed values of all time and frequency domain parameters, as well as the slopes of the stylization line segments in $\text{dB/oct}$.

FIG. 2 – Glottal-flow spectra issued from analytical signal models and their stylization with (dashed lines) and without (solid line) spectral tilt [9].

$|\mathcal{F}(U_g^p)(s)| = \frac{1}{\sqrt{2\pi N}} \sum_{j=1}^{n+1} e^{-\frac{(j-1)(j-1)}{2N\Delta t}} U_g^p(j, \Delta t)|$

FIG. 3 – Glottal-flow derivative and separation point position evolution in time (a) and glottal-flow spectrum (c) issued from numerical simulations with the default values of the model control parameters. Measured acoustic parameters : $F_0 = 93$ Hz, $A_g = 0.04$ dB, $F_g = 58.2$ Hz, $Q_g = 0.47$ dB, $F_c = 1350$ Hz, open quotient $Q_g = 0.56$, return phase $R_m = 0.049$ and asymmetry coefficient $\alpha_m = 0.66$. 

3. Results

Figure 3 shows the output obtained when simulating vocal-fold oscillations with the Niels Louw model at the default values of its control parameters. The spectrum is stylized with 3 linear segments with $+0.5 \text{dB/oct}$, $-9.9 \text{dB/oct}$ and $-7.5 \text{dB/oct}$ slope. The source spectrum resulting from the simulation with a simple mechanical model presents the general form in the spectral domain shared by the various signal glottal-flow models. The accentuation of the profile undulations succeeding the main peak (compare Figures 2 and 3) is originated by the windowing (or truncation) of the sampled signal.

Notice that even if the separation point position changes in a very small fraction $\Delta t$, of the glottal cycle, its temporal evolution produces a discontinuity in the derivative of glottal volume velocity. This discontinuity is not prescribed in glottal-flow signal models but does not prevent acoustic parameter computation. Nevertheless, when $\Delta t$ increases beyond a certain threshold, this discontinuity may introduce a dip in the spectrum, as it is the case for some low $k_c$ values. Such dips may in turn alter the measurement of the spectral tilt, which might become spuriously positive (see Figure 4). In this case, one must turn to the value of the medium slope for an indicative value of the spectral tilt.

Let us now consider the numerical experiments in which the model control parameters are varied.

Note that these experiments are performed without coupling to the vocal tract, in order to concentrate on the effects of the variation of the source control parameters. We recall that the acoustic feedback of the vocal tract does affect the glottogram shape, producing formant ripples in the glottal-flow derivative. To briefly illustrate the effect of vocal tract coupling on the spectrum for vowel [a], see Figure 5. 

Figure 6 shows the sensitivity of the three stylization slopes of the glottal-flow spectrum to the variation of the vibrating mass $m$ - the remaining parameters at their (constant) typical
values. This parameter has a considerable effect on the spectral tilt, while the initial slope remains almost constant. The difference between the medium and final slope values justifies a three-segment stylization of the glottal flow. In fact, rarely has the stylization been possible with only two segments. The spectral tilt is also sensitive to variations in vocal-fold tension $k$, and vocal-fold stiffness $k_c$ (for $k_c \lesssim 20 \, N/m$).

On the contrary, subglottal pressure seems to have little effect on acoustic parameters (including the spectral tilt) with the only exception of the sound pressure level (SPL) parameters (amplitude of voicing and speed of closure). These results are shown in Figure 7.

Wider variations in spectral tilt are attainable in the high-frequency regimes, i.e. when $m/k$ is sufficiently low. Figure 8 shows the simultaneous variation of vocal-fold mass and tension, with the corresponding values of fundamental frequency and spectral tilt. For these parameter values, vocal folds vibrate with a certain amount of glottal leakage (the transglottal air flow does not reach zero during the quasi-closed phase).

Numerical experiments also show that the evolution of $F_m$ and $F_g$ need not be correlated. With the regression-type method used in this work to determine $F_g$, the correlation between $F_m$ and $F_g$ will depend on the physical parameters subjected to variation. For instance, when the vibrating mass is varied, $F_g$ remains a good approximation of $F_m$ (see Figure 9) - with the exception of a few mass values (around 0.004 g) for which the stylization is performed with two (rather than three) line segments according to the procedure described in the previous section. On the other hand, the behavior of $F_m$ and $F_g$ is clearly different as the vocal-fold tension is varied, as shown in Figure 9.

4. Conclusions

We have proposed a method to compute frequency-domain acoustic parameters for glottal-flow spectra computed by means of numerical simulations with an up-to-date two-mass model. The general form of the numerical glottal-flow spectrum coincides with the one prescribed by analytical signal models and conforms to a stylization with three line segments. Vocal tract feedback produces the expected effect on the spectrum and does
not necessarily prevent stylization.

The sensitivity of the features of the glottal-flow spectrum to the variation of source control parameters has been examined in simulations excluding vocal tract feedback. We have observed that the control parameters having an important effect on spectral tilt are the effective vibrating mass and vocal-fold tension. Below a critical value, the variation of vocal-fold stiffness is also shown to produce important variations of the spectral tilt.

Numerical experiments have also shown that the mobility of the separation point introduced in two-mass models in 1998 has the property of introducing a dip in the glottal flow spectrum, which may drastically change the spectral tilt value if a regression-type method is used to compute the stylization slopes. Since the presence of dips in the low- and high-frequency regions of the spectrum is known to be hardly perceived by the listener, it would be convenient to develop a method avoiding such spurious results (appearing close to the bounds of the regions of allowed oscillations). There being still no consensus on how to measure the features of the glottogram spectrum in the literature [13, 14], this is a point which should be kept in mind in future research. It is possible to conclude that the three-line stylization remains possible throughout the inspected ranges: the spectral model is robust under source control parameter variation.

5. References