# Perceptual equivalence of the Liljencrants-Fant and linear-filter glottal flow models

Olivier Perrotin, Lionel Feugère, Christophe d'Alessandro

JASA ARTICLE

# Perceptual equivalence of the Liljencrants–Fant and linear-filter glottal flow models

Olivier Perrotin,[1,a)] Lionel Feugère,[2,b)] and Christophe d'Alessandro[3,c)]
[1]*Université Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, F-38000 Grenoble, France*
[2]*Natural Resources Institute, University of Greenwich, Chatham, Kent ME4 4TB, United Kingdom*
[3]*Sorbonne Université, CNRS, Institut Jean Le Rond d'Alembert, Équipe Lutheries-Acoustique-Musique, F-75005 Paris, France*

**ABSTRACT:**
Speech glottal flow has been predominantly described in the time-domain in past decades, the Liljencrants–Fant (LF) model being the most widely used in speech analysis and synthesis, despite its computational complexity. The causal/anti-causal linear model ($LF_{CALM}$) was later introduced as a digital filter implementation of LF, a mixed-phase spectral model including both anti-causal and causal filters to model the vocal-fold open and closed phases, respectively. To further simplify computation, a causal linear model ($LF_{LM}$) describes the glottal flow with a fully causal set of filters. After expressing these three models under a single analytic formulation, we assessed here their perceptual consistency, when driven by a single parameter $R_d$ related to voice quality. All possible paired combinations of signals generated using six $R_d$ levels for each model were presented to subjects who were asked whether the two signals in each pair differed. Model pairs $LF_{LM}$–$LF_{CALM}$ were judged similar when sharing the same $R_d$ value, and LF was considered the same as $LF_{LM}$ and $LF_{CALM}$ given a consistent shift in $R_d$. Overall, the similarity between these models encourages the use of the simpler and more computationally efficient models $LF_{CALM}$ and $LF_{LM}$ in speech synthesis applications.
 https://doi.org/10.1121/10.0005879
(Received 28 April 2021; revised 2 July 2021; accepted 22 July 2021; published online 19 August 2021)
[Editor: Paavo Alku] Pages: 1273–1285

## I. INTRODUCTION

The acoustic theory of speech production formalised by Fant (1960) assumes independence and linearity between the airflow modulated in the glottis by the vibration of the vocal folds, called glottal flow, and the resonance effect of the vocal tract that shapes the glottal flow into a speech signal. The linear acoustic theory offers a somewhat simplified view of the physics of speech production, but it is still a very effective and widely used representation of voice signals for speech processing applications (e.g., speech coding, synthesis, parameterization) and acoustic phonetics analyses. In this theory, vocal tract resonances introduce spectral formants and anti-formants (maxima and minima of the spectral envelope) that characterise speech sounds. Vocal tract formants are themselves often associated with linear filters: series or parallel branches of second order resonant sections in formant synthesisers; auto-regressive filter models in linear prediction. In early applications, the voice source component was also considered as a low-pass filter, the so-called glottal formant. The transmission line analog proposed by Fant (1960) used a four-pole model subsequently simplified in a two-pole model in linear prediction of speech by Markel and Gray (1982). Note that this glottal formant is not related to a physical resonance but describes the spectrum of the glottal pulse, modelled as the impulse response of the low-pass filter. However, glottal filter impulse responses poorly match glottal flow waveforms obtained by inverse filtering or by indirect measurements like electroglottography. This has led to the proposition of a multiplicity of glottal flow models (GFMs) defined in the time-domain by analytic and parametric formulations of the glottal flow waveform and its derivative: Rosenberg (1971) (Rosenberg model); Hedelin (1984), Fujisaki and Ljungqvist (1986), and Klatt and Klatt (1990) (KLGLOTT88 model); Fant *et al.* (1985) [Liljencrants–Fant (LF) model]; Veldhuis (1998) (R++ model). These widely used models adopt various mathematical functions to describe the glottal flow oscillation, yet Doval *et al.* (2006) showed that the Rosenberg, KLGLOTT88, LF, and R++ models can be grouped under one general expression that is parameterized by a common set of five parameters. Variations of these parameters are closely related to voice quality perception (e.g., breathiness, tenseness, vocal force), that strongly motivates the use of GFM in expressive speech related research. This includes analysis of emotion in speech [Gobl and Ní Chasaide (2003), Patel *et al.* (2011), and Ní Chasaide *et al.* (2013): LF model; Burkhardt and Sendlmeier (2000): KLSYNTH88 model]; analysis-resynthesis schemes for voice modification [Childers (1995), Cabral *et al.* (2014),

a)Electronic mail: olivier.perrotin@grenoble-inp.fr, ORCID: 0000-0002-9909-6078.
b)ORCID: 0000-0003-0883-5224.
c)ORCID: 0000-0002-2629-8752.

and Degottex *et al.* (2013): LF model]; or expressive text-to-speech synthesis [Raitio *et al.* (2013), Airaksinen *et al.* (2016), and Juvela *et al.* (2019): LF model]. This list is not exhaustive; however, LF has been the most widely adopted model for analysis and synthesis of speech signals.

The main limitation of the LF model is its computational complexity. It requires solving implicit equations that can only be performed with numerical approaches. This model is not suitable for applications where computational complexity is a constraint, such as real-time speech or singing synthesis. Also, spectral glottal flow models are desirable because voice quality is often described in spectral terms (e.g., voice spectral tilt, brightness, tenseness): spectral parameters are closer to perception than time-domain parameters. It is therefore interesting to investigate the apparent discrepancy between GFM like LF and filter impulse-response models. Along this line, Doval *et al.* (2006) highlighted that LF and the other time-domain models under study have a simple magnitude representation in the frequency-domain that can be modelled with a third order filter, as also noted by Childers and Lee (1991). This has led to the proposal of new models: the causal/anti-causal linear model ($\mathrm{LF_{CALM}}$) by Doval *et al.* (2003), followed by an all-causal linear model ($\mathrm{LF_{LM}}$) used in the Cantor Digitalis singing synthesiser (Feugère *et al.*, 2017), which both gradually simplify the computation of the glottal flow by using digital filters instead of analytic functions, thus enabling a precision-complexity trade-off, LF being the most precise and $\mathrm{LF_{LM}}$ the simplest. While we will show in Sec. II that the simplification operating on $\mathrm{LF_{CALM}}$ and $\mathrm{LF_{LM}}$ can substantially modify the glottal flow waveform, it is not clear if this affects their auditory perception. The aim of this paper is threefold. Section II studies the three models LF, $\mathrm{LF_{CALM}}$, and $\mathrm{LF_{LM}}$ in terms of linear filters. Formulations for impulse responses are derived, and differences between the models are investigated. After this objective and analytic comparison, subjective experiments are conducted in Sec. III for assessing the perceptual equivalence of the three models. Armed with analytic formulations and perceptual analyses, the discussion in Sec. IV summarises the results obtained: linear-filter formulations equivalent to the LF model are able to account for both the observed glottal formant and glottal flow waveforms.

## II. LINEAR-FILTER FORMULATION OF GLOTTAL FLOW MODELS

### A. Glottal flow model parameters: LF and $R_d$

All GFMs attempt to describe a vocal-fold vibration period in time-domain (see Fig. 1). Three phases are considered: the opening phase (lung pressure forces the vocal folds to spread, and an increasing air flow passes through the glottis); the closing phase (the elasticity of the vocal folds takes over, closing the air passage); the closed phase (the airflow is blocked). Then the lung pressure increases again, and a new opening phase follows. This cycle can be represented by five parameters (Doval *et al.*, 2006): the cycle period $T_0$ or fundamental frequency $F_0 = 1/T_0$; the cycle amplitude, generally represented by $E$, the maximum of the absolute value of the glottal flow derivative (GFD) (i.e., the negative peak at the glottal closure instant has amplitude $-E$); the open quotient $O_q$, the ratio of the open phase duration $T_e$ over the period $T_0$; the asymmetry coefficient $\alpha_m$, the ratio of the opening phase duration $T_p$ over the open phase duration; and $T_a$, the closing time duration (Fig. 1). Period $T_0$ and amplitude $E$ change the time and amplitude scales of the glottal flow. The three other parameters change the shape of the glottal flow and account for the voice timbre or quality. Empirically, Fant *et al.* (1994) established that the perceptual effect of the shape parameters $O_q$, $\alpha_m$, and $T_a$ can be gathered into a unique high-level parameter called $R_e$ initially, $R_d$ afterward (Fant, 1995) (see Appendix A). Typical values of $R_d$ range from 0.4 (short open phase, strong asymmetry of the glottal flow leading to a tense voice) to 2.7

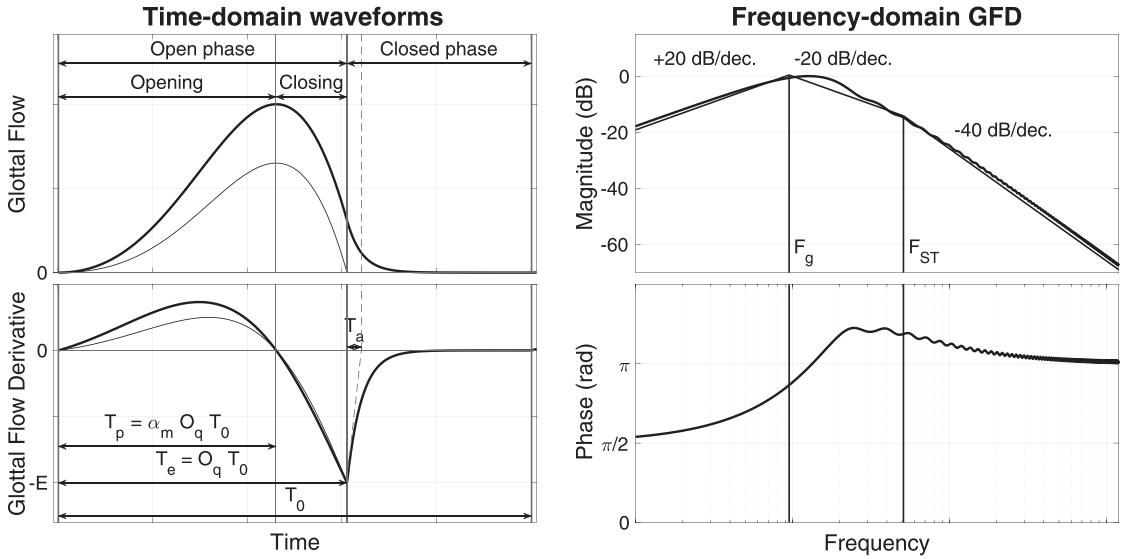


FIG. 1. Left: Temporal parameters of the LF model on the glottal flow (top) and its derivative (bottom). Right: Spectrum magnitude (top) and phase (bottom) of the GFD.

https://doi.org/10.1121/10.0005879

(long open phase, symmetry of the glottal flow providing a relaxed voice). Note that the time-domain NAQ-coefficient proposed later (Alku *et al.*, 2002) is proportional to $R_d$. $R_d$ will be used as a control parameter below.

### B. Glottal formant, $LF_{CALM}$, and $LF_{LM}$

Radiated sound pressure outside the vocal tract can be approximated by the derivative of the speech flow measured at the lips. For this reason, the glottal flow derivative is often preferred to the glottal flow for analysis purposes. The spectrum of the glottal flow derivative shows a marked spectral peak, the glottal formant. Figure 1 displays on the right the magnitude spectrum of the glottal flow derivative computed with the LF model, superimposed on a third order filter approximation. Two poles form the glottal formant, a low-frequency resonance with centre frequency $F_g$ that is directly related to the oscillation of the open phase of the vocal folds. The remaining pole is an extra attenuation with cut-off frequency $F_{ST}$, called spectral tilt, that is responsible for the smoothness of the closed phase of the vocal folds. Phase analysis has shown that this third order approximation is a mixed-phase model (Gardner and Rao, 1997), allowing it to represent the open phase of the LF model as a second order filter response (damped sinusoid) that evolves toward negative time, while the closed phase resembles the response of a first order filter (decreasing exponential) that evolves toward positive time (bottom-left of Fig. 1). Following this analysis, the causal/anti-causal linear model of the glottal flow ($LF_{CALM}$) has been proposed by Doval *et al.* (2003) to generate a glottal derivative waveform by filtering a pulse train with the mixed-phase third order filter. The $LF_{CALM}$ is a simple formulation reproducing the dual relations between time-domain parameters and spectral shape (Gobl *et al.*, 2018; Henrich *et al.*, 2001). A real-time implementation of the model called RT-CALM was then derived by d'Alessandro *et al.* (2006). The mixed-phase characteristic of the glottal flow has been exploited for the estimation of the glottal flow from speech signals (Bozkurt *et al.*, 2005; Drugman *et al.*, 2011; Hézard *et al.*, 2013). The glottal formant can also be represented by causal filters, following Klatt (1980) and Holmes (1983), but at the expense of some distortion in the phase spectrum compared to the LF model. A formulation of this causal linear voice source model $LF_{LM}$ has been proposed and used for real-time voice synthesis and voice source analysis (Feugère *et al.*, 2017; McLoughlin *et al.*, 2020; Perrotin and McLoughlin, 2019, 2020). The perceptual effect of this phase difference is studied in Sec. III.

To summarise, the LF model that is widely accepted as a precise time-domain GFM has been simplified by a frequency-domain representation that uses a mixed-phase third order filter called $LF_{CALM}$. To go further in reducing computation complexity, an all-causal linear model ($LF_{LM}$) has been recently formulated.

All three GFMs are defined in terms of their open and closed phase, described separately in Secs. II C and II D. For this reason, we define glottal opening instants (GOIs) that mark the beginning of each open phase and are spaced by a duration of $T_0$ and glottal closure instants (GCIs) marking the beginning of each closed phase. GOIs and GCIs are spaced by a duration of $O_qT_0$.

### C. Modeling the open phase

#### 1. General formulation of the open phase

Let us define the impulse response of a truncated second order filter, whose generic formulation is

$$\begin{cases} h_T(t) = G_n e^{a_n t} \sin(b_n t + \phi_n) & \text{if } t \in \mathcal{D} \\ h_T(t) = 0 & \text{elsewhere.} \end{cases} \tag{1}$$

If $h_T$ is anti-causal, $T < 0$ is the instant of truncation, $\mathcal{D} = [T, 0]$, and $a_n > 0$. Its causal counterpart is defined for $T > 0$, $\mathcal{D} = [0, T]$, and $a_n < 0$. It can be shown that the open phase definitions of the three GFMs under study can be formulated with respect to Eq. (1) by setting appropriately the $G_n$, $a_n$, $b_n$, $\phi_n$, and $T$ parameters (index $n$ is subsequently replaced by the name of the model in consideration: LF, CALM, or LM). In their original formulations, LF is defined as a continuous time-domain function, while $LF_{CALM}$ and $LF_{LM}$ are defined as digital filters ($Z$-domain). For the sake of generalisation, all expressions are given below as equivalent continuous representations (time and Laplace domains), and derivation details from the original papers' formulations are given in Appendixes B, C, and D.

*a. LF.* The LF model (Fant *et al.*, 1985) is defined by an analytic function in the time-domain relative to the GOI and can be interpreted as an unstable, divergent, and truncated causal filter. However, re-parameterization with $O_q$ and $\alpha_m$ and setting the time origin at the GCI (see Appendix B) allow us to express LF as an anti-causal filter truncated at $T_{LF} = -O_qT_0$, matching Eq. (1). The equations below give the resulting waveform analytic expression and its Laplace transform:

$$\begin{cases} h_{LF_{open}}(t) = \dfrac{-E}{\sin\left(\dfrac{\pi}{\alpha_m}\right)} e^{a_{LF}t} \sin\left(\dfrac{\pi}{\alpha_m O_q T_0} t + \dfrac{\pi}{\alpha_m}\right), \quad t \in [-O_qT_0, 0] \\ H_{LF_{open}}(s) = \displaystyle\int_{T_{LF}}^{0} h_{LF_{open}}(t) e^{-st} dt = \dfrac{G_{LF} b_{LF}(e^{-sT_{LF}} - 1) + Es}{(a_{LF} - s)^2 + b_{LF}^2}. \end{cases} \tag{2}$$

One can now identify from the top equation the values of parameters $G_{LF}$, $a_{LF}$, $b_{LF}$, $\phi_{LF}$, and $T_{LF}$ that are summarised in Table I. $a_{LF}$ is the open phase damping coefficient. It is set so that the airflow of a period is zero and results from an implicit equation (see Appendix B).

*b. $LF_{CALM}$.* The causal/anti-causal linear model uses a second order anti-causal and truncated bandpass filter to model the open phase of the glottis (Doval *et al.*, 2003), whose equation and parameters are derived in Appendix C. The time-domain response of $\mathrm{LF_{CALM}}$, truncated at $T_{\mathrm{CALM}} = -O_q T_0$, and the frequency-domain response are given by computing the inverse $Z$-transform and Laplace transform of the filter, respectively,

$$\begin{cases} h_{\mathrm{CALM}_{open}}(t) = -\dfrac{E}{\sin(\pi(1-\alpha_m))} e^{a_{\mathrm{CALM}} t} \sin\left(\dfrac{\pi}{O_q T_0} t + \pi(1-\alpha_m)\right), \quad t \in [-O_q T_0, 0] \\ H_{\mathrm{CALM}_{open}}(s) = \displaystyle\int_{T_{\mathrm{CALM}}}^{0} h_{\mathrm{CALM}_{open}}(t) e^{-st} dt = \dfrac{(1 + e^{(a_{\mathrm{CALM}} - s)T_{\mathrm{CALM}}}) Es}{(a_{\mathrm{CALM}} - s)^2 + b_{\mathrm{CALM}}^2}. \end{cases} \tag{3}$$

We find again the general formulation of Eq. (1), and the $\mathrm{LF_{CALM}}$ parameters are summarised in Table I.

*c. $LF_{LM}$.* The $\mathrm{LF_{LM}}$ model (Feugère *et al.*, 2017) is the causal version of $\mathrm{LF_{CALM}}$ with the difference that the filter is not truncated, since it converges (see Appendix D). The time and frequency responses of $\mathrm{LF_{LM}}$, whose parameters are given in Table I, are again given by computing the inverse $Z$-transform and Laplace transform of the filter,

$$\begin{cases} h_{\mathrm{LM}_{open}}(t) = \dfrac{E}{\sin(\pi(1-\alpha_m))} e^{a_{\mathrm{LM}} t} \sin\left(\dfrac{\pi}{O_q T_0} t - \pi(1-\alpha_m)\right), \quad t > 0 \\ H_{\mathrm{LM}_{open}}(s) = \displaystyle\int_{0}^{\infty} h_{\mathrm{LM}open}(t) e^{-st} dt = \dfrac{Es}{(a_{\mathrm{LM}} - s)^2 + b_{\mathrm{LM}}^2}. \end{cases} \tag{4}$$

### 2. Comparison between the GFM open phases

Figure 2 displays the open phases of LF (blue), $\mathrm{LF_{CALM}}$ (orange), and $\mathrm{LF_{LM}}$ (green) for the glottal flows (top-left), GFDs (bottom-left), and spectrum of the GFD (right), computed with $R_d = 1.84$ and $E = 0.2$. The top-right of Fig. 2 displays similarities between the three models. First, all open phases derive from second order filters, as their respective Laplace transforms $H_{\mathrm{LF}_{open}}$, $H_{\mathrm{CALM}_{open}}$, and $H_{\mathrm{LM}_{open}}$ all show a similar denominator with a complex conjugate pole. This results in $\pm 20$ dB/decade asymptotes. In particular, all Laplace transforms simplify to $E/s$ at high frequencies, resulting in similar asymptotes for the three GFMs. At low frequencies, the asymptotes are shifted between models but only from a few dB.

TABLE I. GFM parameters and implementations.

| | LF | $\mathrm{LF_{CALM}}$ | $\mathrm{LF_{LM}}$ |
|---|---|---|---|
| $b_n$ | $\dfrac{\pi}{\alpha_m O_q T_0}$ | $\dfrac{\pi}{O_q T_0}$ | $\dfrac{\pi}{O_q T_0}$ |
| $a_n$ | $> 0$ | $\dfrac{\pi}{O_q T_0 \tan(\pi(1-\alpha_m))}$ | $\dfrac{-\pi}{O_q T_0 \tan(\pi(1-\alpha_m))}$ |
| $\phi_n$ | $\dfrac{\pi}{\alpha_m}$ | $\pi(1-\alpha_m)$ | $-\pi(1-\alpha_m)$ |
| $G_n$ | $\dfrac{-E}{\sin(\phi_{\mathrm{LF}})}$ | $\dfrac{-E}{\sin(\phi_{\mathrm{CALM}})}$ | $\dfrac{-E}{\sin(\phi_{\mathrm{LM}})}$ |
| $T$ | $-O_q T_0$ | $-O_q T_0$ | $\infty$ |
| | *Open phase* | | |
| *Formulation* | Analytic | Filter | Filter |
| *Causality* | Anti-causal | Anti-causal | Causal |
| *Truncation* | At $-O_q T_0$ | At $-O_q T_0$ | No truncation |
| | *Closed phase* | | |
| *Formulation* | Analytic | Filter | Filter |
| *Causality* | Causal | Causal | Causal |

LF and $\mathrm{LF_{CALM}}$ display two more similarities. First, their anti-causality causes the GFD phase to increase (bottom-right of Fig. 2); second, they are both truncated at $t = -O_q T_0$. The thin dashed curves in the left panels show what would be non-truncated versions of LF and $\mathrm{LF_{CALM}}$. A direct effect of the truncation is the computation of their Laplace transform on the interval $[-O_q T_0, 0]$, which results in the appearance of the term $e^{-sT}$ in $H_{\mathrm{LF}_{open}}$ and $H_{\mathrm{CALM}_{open}}$. This causes the ripples observed in the LF and $\mathrm{LF_{CALM}}$ spectra. The main difference between LF and $\mathrm{LF_{CALM}}$ is that the former is parameterized to be class $C^1$ i.e., with a continuous GFD at the GOI ($-O_q T_0$). This parameterization results in a generic second order filter that is neither low-pass nor bandpass, as shown by the numerator of $H_{\mathrm{LF}_{open}}$. A consequence is the large lobe around the resonance frequency of the GFD magnitude spectrum. Conversely, $\mathrm{LF_{CALM}}$ is parameterized to be a bandpass filter, which allows a reduction of the resonance's lobewidth but cannot suppress it completely because of the effect of truncation. The consequence of the bandpass parameterization is a discontinuous GFD at the glottal opening instant.

Two differences between $\mathrm{LF_{CALM}}$ and $\mathrm{LF_{LM}}$ are also highlighted. The difference of causality is well-displayed by a vertical symmetry in the time-domain and a horizontal

https://doi.org/10.1121/10.0005879

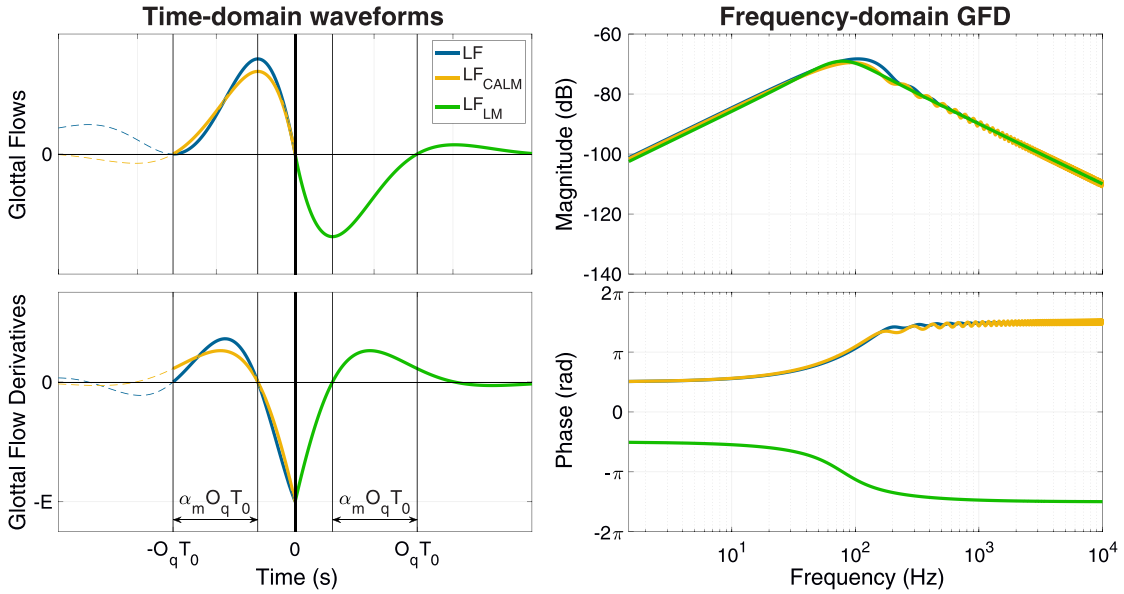


FIG. 2. (Color online) LF (blue), $\mathrm{LF_{CALM}}$ (orange), and $\mathrm{LF_{LM}}$ (green) open phases. Left: Glottal flows and their derivatives. Right: Magnitude and phase spectrum of the glottal flow derivatives.

symmetry of the phase spectrum. Also, because $\mathrm{LF_{LM}}$ converges, it is not truncated at $O_qT_0$. This implies a leak of the period to the next one but also greatly simplifies its implementation. As a result, its spectrum is the exact frequency response of a bandpass filter, with no ripples and no lobe around the resonance centre frequency. Note that the vertical symmetry of the GFD between $\mathrm{LF_{CALM}}$ and $\mathrm{LF_{LM}}$ implies a sign inversion of the glottal flow, but one that the ear is not sensitive to.

### D. Modeling the closed phase

#### 1. Formulation of the closed phases

Definitions of the GFM closed phase fall within two categories (Doval *et al.*, 2006): it is either described in the time-domain by an analytic formulation, as LF, or defined in the frequency-domain with a first order filter, as $\mathrm{LF_{CALM}}$ or $\mathrm{LF_{LM}}$.

*a. LF: Analytic expression.* The closed phase of the LF model, after shifting the glottal closing instant at $t = 0$, is expressed as

$$h_{\mathrm{LF}_{closed}}(t) = \frac{-E}{\epsilon T_a}\left(e^{-\epsilon t} - e^{-\epsilon(T_0 - O_qT_0)}\right), \quad t \in [0, T_0 - O_qT_0], \tag{5}$$

where $\epsilon$ is the closed phase coefficient. It satisfies the continuity of the open and closed phase expressions at the GCI and is obtained from an implicit equation (see Appendix B). Note that because $a_{\mathrm{LF}}$ is computed from $\epsilon$, the shape of the open phase depends on the closed phase, although both phases are defined by distinct analytical expressions.

*b. $\mathrm{LF_{CALM}}$ and $\mathrm{LF_{LM}}$: Filtering.* With $\mathrm{LF_{CALM}}$ and $\mathrm{LF_{LM}}$, the closed phase is modelled by a first order low-pass filter attenuating high frequencies above its cut-off frequency $F_a = 1/(2\pi T_a)$ and called spectral tilt (Doval *et al.*, 2003; Feugère *et al.*, 2017). Filter formulation is given in Appendix C. In these cases, the spectral tilt filter is applied on the full signal and therefore changes the open phase shape.

#### 2. Comparison between the GFM closed phases

Figure 3 displays the three GFM full waveforms, obtained by adding to the open phases of Fig. 2 their respective closed phase contributions while keeping $R_d = 1.84$. Note that this process changes the open phases. The top-right panel shows high similarity between the three GFMs' spectrum magnitudes. The closed phase adds a supplementary $-20$ dB/decade attenuation to all open phase spectra, resulting in a $-40$ dB/decade attenuation at high frequencies. We can also observe an increase in gain in low frequencies for the LF model. This is directly linked to the change of the $a_{\mathrm{LF}}$ parameter. A consequence is the largest amplitudes of the glottal flow and glottal flow derivative for LF.

Looking at the phase spectrum (bottom-right panel), LF and $\mathrm{LF_{CALM}}$ almost overlap, showing a similar effect of the closed phases on their respective phase spectra: it adds a supplementary $-\pi/2$ offset at high frequencies to all phase spectra of the open phases. The spectral tilt filter is displayed in black. This offset introduces an asymmetry between LF and $\mathrm{LF_{CALM}}$ on one side and $\mathrm{LF_{LM}}$ on the other. The addition $-\pi/2$ at high frequencies for all models reduces the phase of LF and $\mathrm{LF_{CALM}}$ from $3\pi/2$ to $\pi$ but also reduces the phase of $\mathrm{LF_{LM}}$ from $-3\pi/2$ to $-2\pi$. This asymmetry is reflected in the shapes of the glottal flow derivatives (bottom-left panel). One can see that $\mathrm{LF_{LM}}$ and $\mathrm{LF_{CALM}}$ are not symmetrical anymore and that the filtering attenuates more the GFD peak near the glottal closure instant for $\mathrm{LF_{LM}}$ than for $\mathrm{LF_{CALM}}$. Finally, it is important to

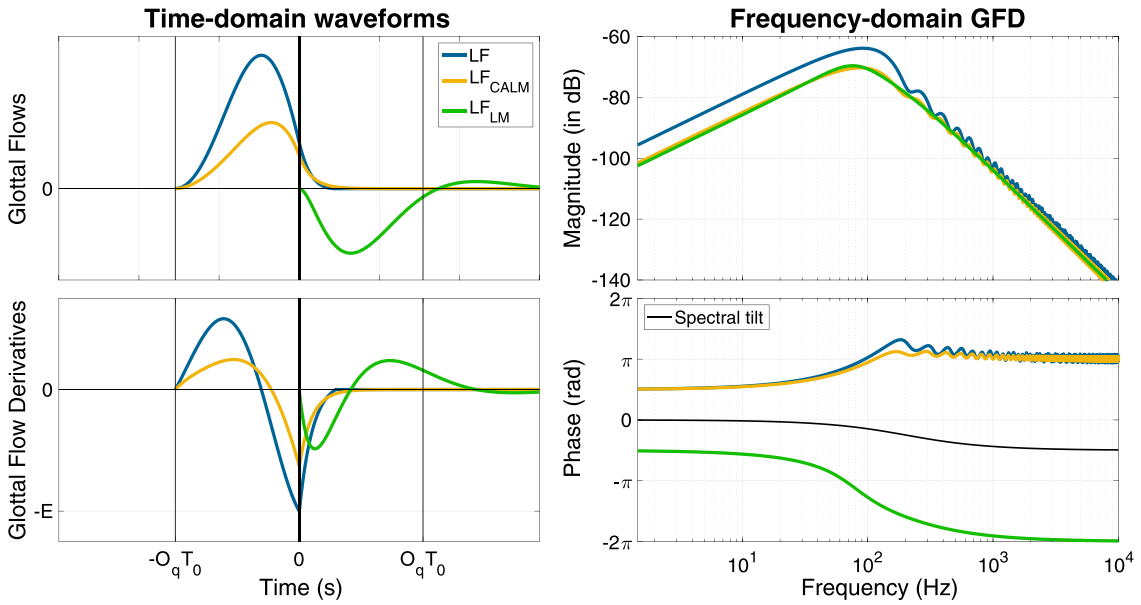


FIG. 3. (Color online) LF (blue), $LF_{CALM}$ (orange), and $LF_{LM}$ (green) waveforms including closed phases. Left: Glottal flows (top) and derivatives (bottom). Right: Magnitude (top) and phase (bottom) spectrum of the glottal flow derivatives.

mention that the spectral tilt filter is not truncated for $LF_{CALM}$ and $LF_{LM}$, and its application results in an infinite response that may overlap with the next period. This appears for high values of $R_d$, as shown in Sec. II F.

### E. Assessment of computational costs

To evaluate the computational efficiency of each GFM, we measured the average time necessary to compute one period of a 1-s stationary signal for each model. The ratio of computation time over the period duration gives the real-time factor. A real-time factor below 1 means that the signal is faster to compute than to play back, so we can listen the signal while it is generated. Inversely, a real-time factor higher than 1 indicates that the signal takes longer to compute than to play back. This experiment was made in the condition of a fine-grain control of the GFM: parameters are calculated for each period. To assess the dependency of the real-time factor on $F_0$ and $R_d$, we generated 564 stationary signals using a combination of the six $R_d$ values described in Sec. III and 94 $F_0$ values, from 70 to 1000 Hz with steps of 10 Hz. All signals were generated on an iMac Intel Core i9, with a 3.6 GHz processor. Figure 4 displays the real-time factors for the three GFMs depending on $F_0$. For each model and $F_0$ value, we computed the mean and standard deviation of the real-time factor across the six $R_d$ values. The means for each model are represented by the thick coloured lines, and the shading around each mean value highlights the $\pm$ standard deviation range around the mean. $LF_{CALM}$ and $LF_{LM}$ are more than 10–100 times faster than LF. This is a direct consequence of the resolution of the implicit equation for the LF model, which is costly. Also, the efficiency of LF decreases with higher $F_0$ because the resolution of the implicit equation requires a constant duration. Therefore, when the period duration decreases, the real-time factor increases, and this dependency between computation efficiency and input parameter is not desirable. Finally, $R_d$ has no effect on the computation time for all three GFMs.

### F. Summary of the model implementation and effect of $R_d$

Table I summarises the implementations of the three GFMs under study. To conclude this section, Fig. 5 shows the effect of $R_d$ on the GFD (top row) and the respective spectra computed on a single period (second and third rows) for the three models [LF (blue), $LF_{CALM}$ (orange), and $LF_{LM}$ (green)]. In the top row, the dashed vertical lines

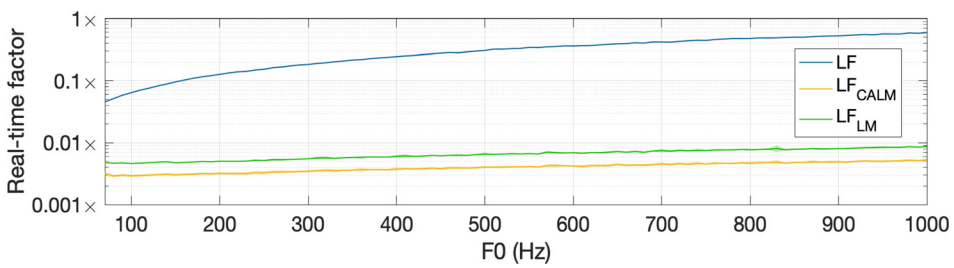


FIG. 4. (Color online) Computational efficiency of each model expressed in real-time factor: LF (blue), $LF_{CALM}$ (orange), and $LF_{LM}$ (green).

https://doi.org/10.1121/10.0005879

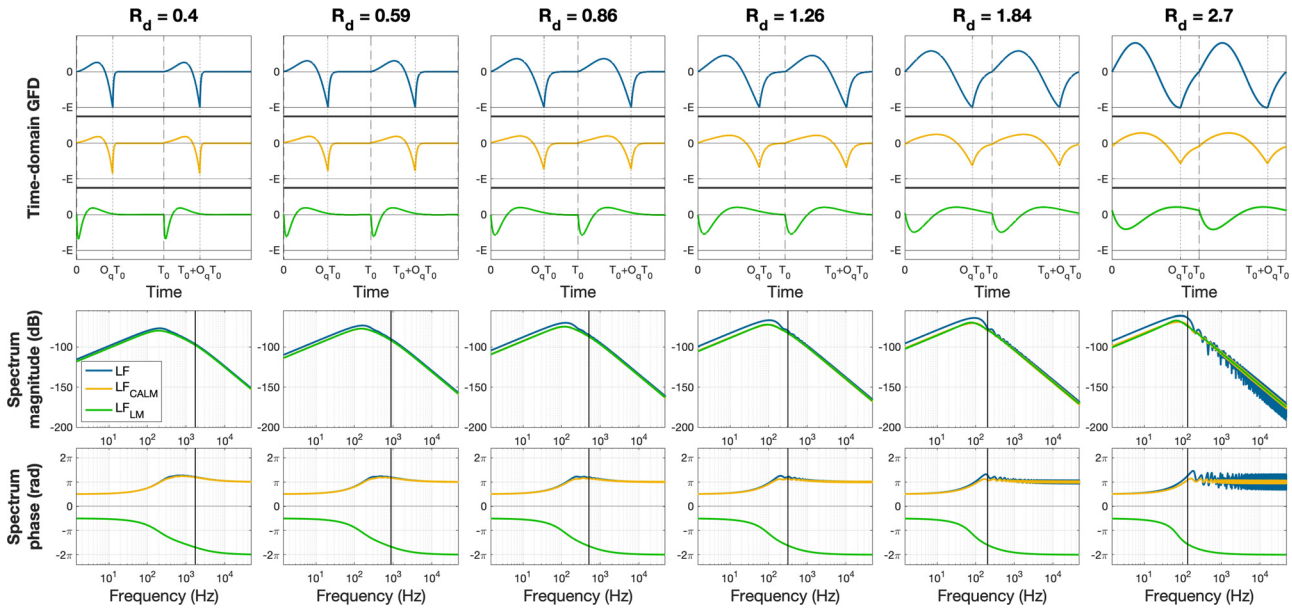


FIG. 5. (Color online) Glottal flow derivatives (top row) and their magnitude (second row) and phase (bottom row) spectra computed with the three models [LF (blue), $LF_{CALM}$ (orange), and $LF_{LM}$ (green)] for a range of $R_d$ values (each column).

represent the GOIs, while the dotted lines show the GCIs. In the second and third row, the vertical line indicates the cut-off frequency of the spectral tilt filter. Globally, $R_d$ has a similar effect on the three GFMs. Looking at the spectrum magnitude, low values of $R_d$ lead to higher centre frequency and bandwidth of the glottal formant and a higher spectral tilt cut-off frequency. These combined effects favour the presence of numerous harmonics that give a sharp GFD closure, close to the shape of an impulse. This is typical for tensed and loud voice, when the vocal folds open and close abruptly. Inversely, high values of $R_d$ lower the centre frequency and bandwidth of the glottal formant as well as the spectral tilt cut-off frequency. It thus emphasises more the first and second harmonics, leading to a more sine-like GFD shape. This is lax/soft voice, when the vocal folds oscillate more symmetrically.

In the first column of Fig. 5, the three GFMs appear very similar for two reasons. First, a low value of $R_d$ leads to a high attenuation coefficient $a_n$ that allows LF and $LF_{CALM}$ to have almost horizontal tangents at GOI. The truncation thus does not introduce an abrupt change of slope on the GFD, which results in a reduction of ripples on the LF and $LF_{CALM}$ spectra. Second, the effect of spectral tilt that introduces an asymmetry between $LF_{CALM}$ and $LF_{LM}$ is small (high cut-off frequency), leading to almost symmetrical $LF_{CALM}$ and $LF_{LM}$ GFDs. Inversely, the three GFM shapes diverge with increasing values of $R_d$. Truncation has stronger effects on LF and $LF_{CALM}$, increasing ripples in their spectrum, and the spectral tilt whose cut-off frequency is closed to the glottal formant position has a strong effect on the GFD shapes. In particular, one can note that the minimum values of $LF_{CALM}$ and $LF_{LM}$ diverge from $-E$ when $R_d$ increases. Moreover, the last column illustrates well the effect of absence of truncation of the spectral tilt filter on $LF_{CALM}$ and $LF_{LM}$. The GFD computed for one period overlaps on the next one, leading to negative (respectively, positive) value of the GFD at the GOI for $LF_{CALM}$ (respectively, $LF_{LM}$).

We have shown that the difference of construction between the three GFMs (formulation, causality, truncation) leads to clear visible differences in the GFD waveforms and spectra. However, their effect on auditory perception is unclear and is assessed in Sec. III.

## III. PERCEPTUAL COMPARISON OF VOICE SOURCE MODELS

### A. Experiment

#### 1. Protocol and task

The aim of the experiment was to assess any perceptual difference between the three GFMs for different values of the $R_d$ parameter. We used for this purpose a two-alternative forced-choice (2AFC) protocol (Kingdom and Prins, 2016), where each subject's task was to listen to paired sounds and to say if they were the same or different, with respect to any distinctive features, whatever their nature (e.g., timbre, level, pitch, etc.). The experiment was divided into three blocks. The first block used synthesised sounds from the GFMs only. The second and third blocks used additional /a/ and /i/ vocal tract models convolved with the GFMs. These two vowels were chosen for their lowest (/i/) and highest (/a/) first formant frequency in order to test a more natural vocal sound than the GFM alone.

For each GFM and following Degottex *et al.* (2013), six values of $R_d$ were chosen equally spaced on a logarithmic scale, leading to three GFMs $\times$ 6 $R_d$ = 18 stimuli per block (the one displayed on Fig. 5). Then for each block, every

combination of pairs of different GMFs was tested ($LF_{LM}$ × $LF_{CALM}$; $LF_{LM}$ × LF; $LF_{CALM}$ × LF; $LF_{CALM}$ × $LF_{LM}$; LF × $LF_{LM}$; LF × $LF_{CALM}$). Finally, 3 vowels × 6 pair combinations × 6 $R_d$ values for the first element of the pair × 6 $R_d$ values for the second element of the pair led to a total of 648 pairs of stimuli to compare.

A computer interface was specially designed for this experiment and programmed in MAX 6.[1] The protocol was identical for all the paired stimuli. To proceed, the subject clicked a button, which launched the playback of two sounds, A and B, separated by 500 ms. The test sounds were ordered randomly and played for each subject only once to keep sessions as short as possible and identical among subjects. The subject had to choose whether the two sounds were identical or different, without any other choice. Each block lasted approximately 10 min, and subjects were especially encouraged to stop and rest between the three blocks with a message displayed automatically. The entire experiment took place in an acoustically insulated and treated room designed for perceptual experiments. Sound was played using a Focusrite (High Wycombe, UK) Scarlett 2i2 audio interface on a Mac OSX and AKG (Los Angeles, CA) K271 headphones. Before the experiment, subjects were trained with a subset of the sound-pair list (GFM convolved to /a/ vocal tract or without vocal tract, three $R_d$ values spread over the full range of possible values).

A group of 18 subjects took part in the experiment (median age of 28 years, from 21 to 54 years old). Among them, 12 subjects worked in the field of sound technologies, and six others had a regular musical practice. An audiogram test was performed for each of the subjects, and none of them reported any known auditory impairment except one who was single-side deaf, but stereo listening was not needed to perform the task. Fourteen subjects were members of the laboratory and participated in the experiment on a voluntary basis without being paid. The four remaining subjects were paid for the experiment.

### 2. Stimuli specification

Stimuli were synthesised at a sampling rate of $F_s = 96$ kHz. A constant fundamental frequency of $F_0 = 110$ Hz and a peak amplitude $E = 0.2$ were chosen. The LF GFDs were generated by using the analytic formulations of Eqs. (2) and (5) and by solving the implicit Eqs. (B3) and (B4). The $LF_{CALM}$ and $LF_{LM}$ GFDs were generated by filtering a pulse train with their respective open and closed phase filters (Appendix E). All signals lasted 0.3 s, a duration longer than a standard spoken syllable but short enough to facilitate recall of the two stimuli for comparison. Fade-in and fade-out amplitudes were applied using half Hanning windows of length $10T_0 = 0.09$ s. Vowels were invariant in time and were applied by filtering the GFM with a bank of five parallel resonant filters corresponding to vowels /i/ and /a/, whose transfer functions are given in Feugère *et al.* (2017). Finally, all stimuli were normalised in dBA.[2]

## B. Results

Results report the proportion of pairs that were judged similar depending on the factors in consideration. In particular, we factorised the six different model pairs into two factors: the *Model*/factor (three levels: LF × $LF_{CALM}$; $LF_{LM}$ × LF; $LF_{CALM}$ × $LF_{LM}$) and the *Order* factor that codes the order of presentation of each pair (two levels). The additional factors are *Vowel* (three levels: source only; /a/; /i/) and $R_d$ (36 levels for all combinations of the six selected values). In the following, we used a single generalised linear model following a binomial distribution to assess the significance of each factor and their interactions for the perception results. The obtained model was subsequently simplified by iteratively removing non-significant interactions between factors provided that, at each simplification step, the current and the simplified models do not significantly differ ($p > 0.05$) (Crawley, 2013). *Post hoc* Pearson's chi-squared tests were run to assess whether proportions obtained for single conditions significantly differ from chance.

Figure 6 shows perceptual experiment responses for all factors and interactions except *Order* (results for both presentation orders of each pair are merged). The top-left panel shows results relative to the $R_d$ factor only. Each square corresponds to the proportion of pairs judged similar for a given couple of $R_d$ values, all *models* and *Vowels* combined. Pairs in black and white were judged similar by 100% and 0% of the subjects, respectively. Scores that fall within the red rectangle on the colour bar do not significantly differ from chance according to the *post hoc* Pearson's chi-squared tests ($p > 0.05$). On the left-hand side of the figure, the top row (columns 2–4) shows the *model* × $R_d$ interaction, with all levels of *Vowel* and *Order* combined; the left column (rows 2–4) shows the *Vowel* × $R_d$ interaction, with all levels of *model* and *Order* combined; the remaining panels show the *Vowel* × *model* × $R_d$ interaction for each level of *Vowel* and *model*, indicated in the top and left margins of the figure. Panels with yellow and green contours are replicated on the right side, with LF put in the abscissa. On top (respectively, bottom) for each $R_d$(LF) value (each column), the distribution of perceived similar $R_d(LF_{CALM})$ [respectively, $R_d(LF_{LM})$] values was obtained and superimposed on the figure, the circles being the medians and the error bars corresponding to 90% of the values around the median. Smaller circles indicate scores below the level of significance (Pearson's chi-squared test). The shaded area links all error bars and represents the space of perceptual equivalence between $R_d$(LF) and $R_d(LF_{CALM})$ (respectively, $R_d(LF_{LM})$).

### 1. Effect of $R_d$ and order

The $R_d$ factor has the strongest effect on results [$R_d$: $\chi^2 = 3620$, degrees of freedom (df) = 35, $p < 0.001$]. The top-left panel of Fig. 6 clearly shows that, over all other factors, pairs with similar values of $R_d$ are strongly perceived as similar, and vice versa. This confirms that $R_d$ has a strong perceptual effect on the synthesis of glottal flow. Presentation order had no influence on similarity judgment

https://doi.org/10.1121/10.0005879

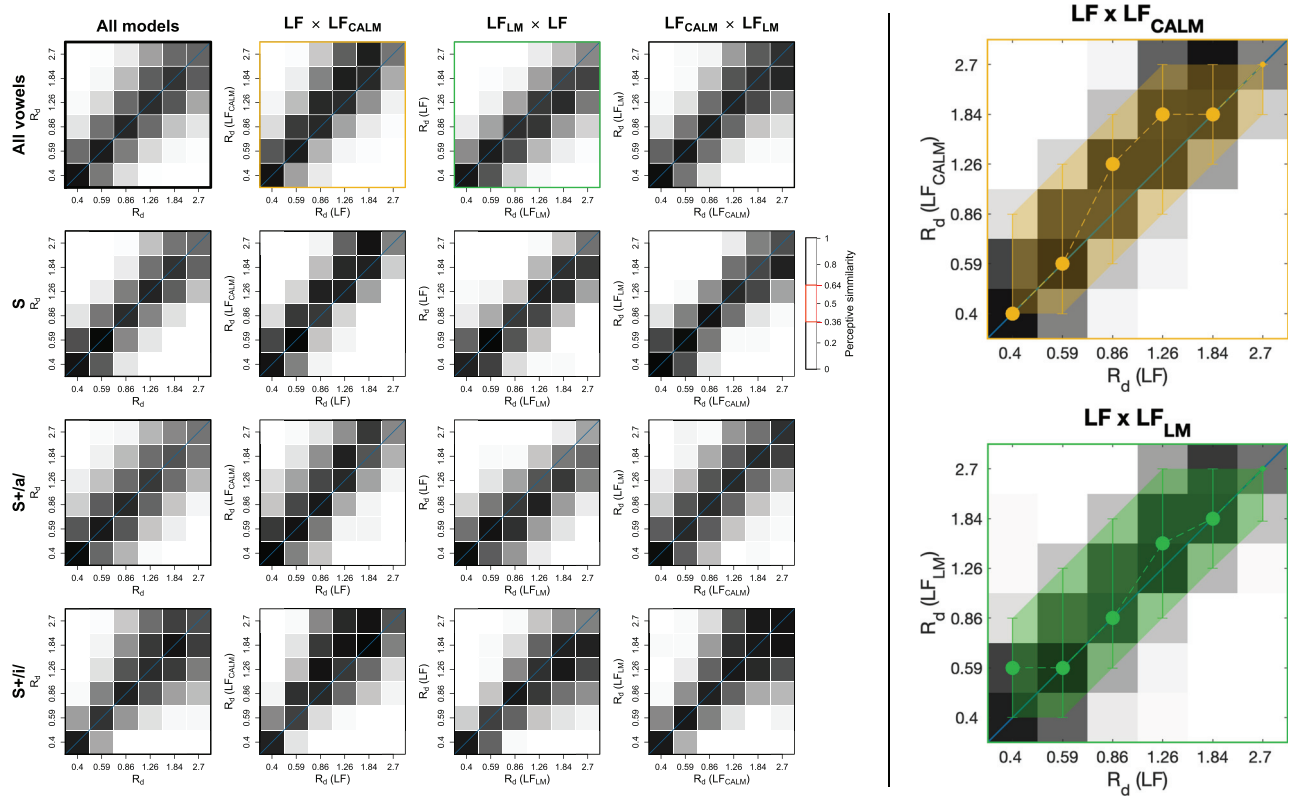


FIG. 6. (Color online) Perceptual experiment answers. Each square panel shows the percentage of pairs judged similar for every couple of $R_d$ values. Black and white squares are stimuli judged similar by 100% and 0% of subjects, respectively. See text for a detailed explanation.

(*Order*: $\chi^2 = 0$, df $= 1$, $p = 0.90$). Therefore, all results displayed in Fig. 6 and detailed below combine the scores of both presentation orders.

### 2. Effect of model

The *model* factor alone has a small and marginally significant effect on subjects' scores (*model*: $\chi^2 = 8.8$, df $= 2$, $p = 0.012$) and therefore demonstrates that the three models are perceptually close to each other. $LF_{CALM}$ and $LF_{LM}$ are judged the most similar models, and LF and $LF_{LM}$ are judged the least similar when all answers are averaged. The subjects' perception seems to reflect the differences between models' construction that are summarised in Table I. $LF_{CALM}$ and $LF_{LM}$ derive from the same filtering process, with the only difference being the causality of the open phase and the truncation of $LF_{CALM}$. Inversely, LF and $LF_{LM}$ differ at almost every point of Table I. While these results average all possible $R_d$ pairs, results depending on $R_d$ follow the significant two-way interaction between *model* and $R_d$ (*model* × $R_d$, $\chi^2 = 486$, df $= 70$, $p < 0.001$). Corresponding results are shown in the top row of the left side of Fig. 6 (columns 2–4). The first observation is that stimuli with similar values of $R_d$ are judged extremely similar (close to 100% similarity), while stimuli with different values of $R_d$ are judged different (0% similarity). One can then note a diagonal asymmetry in the LF × $LF_{CALM}$ and $LF_{LM}$ × LF panels for $R_d$ values higher than 0.86, i.e., when the models start to differ the most (Fig. 5). In particular, subjects judged LF and $LF_{CALM}$ similar mostly when $R_d(LF_{CALM})$ was greater than or equal to $R_d(LF)$. Similarly, LF and $LF_{LM}$ were mostly judged similar when $R_d(LF_{LM})$ was greater than to or equal to $R_d(LF)$. Conversely, $LF_{CALM}$ and $LF_{LM}$ were judged the most similar when they shared the same $R_d$ value, picturing more symmetric results (top-right panel of the left side of Fig. 6).

The right side of Fig. 6 summarises this asymmetry between LF and the other models. Recall that these panels are replicates of the one with yellow and green contours from the left-hand side, but with LF put in the abscissa for both plots. For each $R_d(LF)$, medians of corresponding distributions of perceived similar $R_d(LF_{CALM})$ [respectively, $R_d(LF_{LM})$] are all on or above the diagonal. Also, the spread of each distribution represented by the error bars (90% of the values around the median) and emphasised by the shaded areas clearly displays asymmetrical spaces of perceptual equivalence between $R_d(LF)$ and $R_d(LF_{CALM})$ [respectively, $R_d(LF_{LM})$] that are again above the diagonal, with $R_d(LF_{CALM})$ [respectively, $R_d(LF_{LM})$] mostly equal to or greater than corresponding $R_d(LF)$.

### 3. Effect of vowel

The effect of vowels (*Vowel*: $\chi^2 = 17.5$, df $= 2$, $p < 0.001$) supports that GFDs presented alone were significantly judged less similar than when they were passed through a vowel, the vowel /i/ giving the highest similarity results.

Therefore, the introduction of resonances in the signal mitigates the perception of the glottal source timbre. Moreover, the glottal formant $F_g$ evolves within the range [64, 121] Hz for the chosen values of $R_d$ for all models. The vowel /i/, having its first formant resonance the closest to $F_g$, could mask the effect of $R_d$ variation, leading to sources judged more similar with /i/ rather than vowel /a/.

Also, a significant two-way interaction with $R_d$ is present (*Vowel* × $R_d$: $\chi^2 = 302$, df = 70, $p < 0.001$) as shown in the left column of Fig. 6, rows 2–4. Stimuli presented with the source only show similarity concentrated around the diagonal. When presented with the vowel /i/, the similarity spreads across adjacent $R_d$ values for high $R_d$. This corresponds to $F_g$ and $F_{ST}$ values that are around 100 Hz, close to the first formant frequency of vowel /i/ (215 Hz). Conversely, for the /a/ vowel, it seems that stimuli with high $R_d$ value were neither clearly perceived as similar nor dissimilar. In this case, the first formant frequency (700 Hz) is far above the $F_g$ and $F_{ST}$ ranges. A possibility is that subjects either focused on the low or high frequency parts of the signal, the former hearing the source differences and the latter focusing on the /a/ resonance.

### 4. Remaining interactions

No significant three-way interaction between *Vowel* and *model* and $R_d$ was detected. It can be seen in Fig. 6 that the trend previously observed in the top row and left column (two-way interactions) applies to the remaining plots. Statistical analysis did not reveal a significant *Vowel* × *model* interaction, showing that the perception of differences between models is relatively independent from the addition of a vocal tract. Although it would be necessary to cover a larger number of vocal tract configurations, this finding encourages the hypothesis that the choice of the glottal flow can be made independently from the behavior of the vocal tract. Finally, two-way interactions *Order* × $R_d$ and *Order* × *model* result from the asymmetry of the *model* levels ($\chi^2 = 98$, df = 35, $p < 0.001$; $\chi^2 = 7.6$, df = 2, $p = 0.022$, respectively). The top row of Fig. 6 showed an asymmetry between LF and $LF_{CALM}$ and between $LF_{LM}$ and LF. When considering the order of presentation as a factor, e.g., distinguishing LF × $LF_{CALM}$ vs $LF_{CALM}$ × LF, the asymmetry of $LF_{CALM}$ × LF results is reversed compared to LF × $LF_{CALM}$, hence the two-way interaction.

## IV. DISCUSSION AND CONCLUSION

In this study, the LF model is reformulated in terms of linear filters. This formulation reconciles the apparent discrepancy between time-domain GFM and spectral voice source models. It allows for quantitative spectral interpretation of the LF model parameters because the correspondence between time-domain and spectral parameters can be analytically computed. This unifies Fant's views on the voice source: the key point is the interpretation of the LF GFM [in Fant *et al.* (1985)] as a mixed phase system and not as a simple resonant filter [as in Fant (1960)]. The joint variation of the waveform and glottal formant as a function of $R_d$ can be computed for voice quality analysis and synthesis. As a rule of thumb, increasing $R_d$ corresponds to lowering the glottal formant centre frequency (often referred to as the "voicing bar" in wideband spectrogram reading) and increasing the spectral tilt toward lower frequencies (the right-hand "skirt" of the glottal formant).

Following the proposal of glottal flow models that attempt to reduce the computational complexity of LF, namely $LF_{CALM}$ and $LF_{LM}$, we sought to assess the perceptual consistency of these models. We first showed that even though LF is defined from an analytic expression and $LF_{CALM}$ and $LF_{LM}$ from digital filters, they can all be expressed by the same analytic function, with their own set of parameters. In terms of construction, LF and $LF_{CALM}$ have anti-causal and truncated open phases, while $LF_{LM}$ has a causal and non-truncated open phase. The three GFM closed phases are causal.

Perceptual pairwise-comparison of these models parameterized with various levels of $R_d$ using a same-different forced-choice paradigm on short stationary signals shows that all models are perceived similarly, in that they share the same $R_d$ parameterization with a possible offset. In particular, $LF_{LM}$ and $LF_{CALM}$ are perceived similarly with the same $R_d$, while LF is perceived similarly as $LF_{CALM}$ and $LF_{LM}$ when LF has a smaller $R_d$ value. Investigation seems to show that this shift in perception relates more to the truncation of the glottal flow open phase than to a difference of causality. Nevertheless, this needs to be confirmed in further experiments. Finally, we showed that the addition of vocal tract effect with low vocalic formants increases the perception of similar waveforms when $R_d$ varies slightly between two waveforms. If the high dissimilarity between waveforms (Fig. 3) has favoured the use of LF for precise analysis of the glottal flow (i.e., time-domain analyses), the perceptual consistency between models encourages the use of $LF_{CALM}$ and $LF_{LM}$ as simpler models than LF for speech synthesis applications and for spectral analyses of the voice source and voice quality.

## ACKNOWLEDGMENTS

Part of this work has been done in the framework of the Agence Nationale de la Recherche, through the ChaNTeR and GEPETO Projects (ANR-13-CORD-0011, 2014–2017, ANR-19-CE28-0018, 2019–2023) and "Investissements d'avenir" programs ANR-15-IDEX-02 and ANR-11-LABX-0025-01. The authors are indebted to Professor Boris Doval for his help in the development of the model calculations.

## APPENDIX A: HIGH- TO LOW-LEVEL GLOTTAL PARAMETERS

Fant (1995) derived a unique high-level parameter $R_d$ to control all low-level parameters $O_q$, $\alpha_m$, and $T_a$. He first defined intermediate parameters $R_a$, $R_k$, and $R_g$ from which are derived the low-level parameters,

https://doi.org/10.1121/10.0005879

$$\begin{cases} R_a = (-1+4.8R_d)/100 \\ R_k = (22.4+11.8R_d)/100 \\ R_g = \dfrac{R_k(0.5+1.2R_k)}{0.44R_d - 4R_a(0.5+1.2R_k)} \end{cases} \Rightarrow \begin{cases} O_q = \dfrac{1+R_k}{2R_g} \\ \alpha_m = \dfrac{1}{1+R_k} \\ T_a = R_a T_0. \end{cases} \tag{A1}$$

## APPENDIX B: DERIVATION OF LF

LF is defined in the time-domain by an analytic function (Fant *et al.*, 1985). After re-parameterization with $O_q$ and $\alpha_m$, Doval *et al.* (2006) expressed the open phase of the glottal flow derivative as

$$x_{\mathrm{LF}_{open}}(t) = -\frac{Ee^{-a_{\mathrm{LF}}O_qT_0}}{\sin\left(\frac{\pi}{\alpha_m}\right)} e^{a_{\mathrm{LF}}t} \sin\left(\frac{\pi}{\alpha_m O_q T_0} t\right), \quad t \in [0, O_qT_0]. \tag{B1}$$

Setting the time origin at the glottal closure instant allows us to express LF as an anti-causal filter truncated at $T_{\mathrm{LF}} = -O_qT_0$. This is simply done by defining $h_{\mathrm{LF}_{open}}(t) = x_{\mathrm{LF}_{open}}(t + O_qT_0)$,

$$h_{\mathrm{LF}_{open}}(t) = \frac{-E}{\sin\left(\frac{\pi}{\alpha_m}\right)} e^{a_{\mathrm{LF}}t} \sin\left(\frac{\pi}{\alpha_m O_q T_0} t + \frac{\pi}{\alpha_m}\right), \quad t \in [-O_qT_0, 0]. \tag{B2}$$

Also, if we note $X_{\mathrm{LF}_{open}}$ the Laplace transform of the original formulation given by Eq. (B1), then the time shift operated between $h_{\mathrm{LF}_{open}}$ and $x_{\mathrm{LF}_{open}}$ is translated as $X_{\mathrm{LF}_{open}}(s) = H_{\mathrm{LF}_{open}}(s)e^{-sO_qT_0}$. This linear phase shift does not have any effect on the timbre of the source and is ignored in this paper.

$a_{\mathrm{LF}}$ is the open phase damping coefficient. It is set so that the airflow of a period is zero and thus also depends on the closed phase coefficient $\epsilon$ [Eq. (5)]. The latter satisfies the continuity of the open and closed phase expressions at the GCI from the implicit equation

$$1 - e^{-\epsilon(T_0 - OqT_0)} = \epsilon T_a. \tag{B3}$$

Given the expression of the closed phase, $a_{\mathrm{LF}}$ is calculated so that the integral of the glottal flow derivative is null on a period, leading to the implicit equation

$$\frac{1}{a_{\mathrm{LF}}^2 + (\pi/(\alpha_m O_q T_0))^2} \times \left( e^{-a_{\mathrm{LF}}O_qT_0} \frac{\pi/(\alpha_m O_q T_0)}{\sin(\pi/\alpha_m)} + a_{\mathrm{LF}} - \frac{\pi/(\alpha_m O_q T_0)}{\tan(\pi/\alpha_m)} \right) = \frac{T_0 - O_qT_0}{e^{\epsilon(T_0 - O_qT_0)} - 1} - \frac{1}{\epsilon}. \tag{B4}$$

Both implicit equations are resolved numerically.

## APPENDIX C: DERIVATION OF $\mathrm{LF}_{\mathrm{CALM}}$

The $\mathrm{LF}_{\mathrm{CALM}}$ open phase anti-causal filter is defined in the $Z$-domain by Doval *et al.* (2003) as

$$H_{\mathrm{CALM}_{open}}(z) = \frac{b_1 z + b_2 z^2}{1 + a_1 z + a_2 z^2}. \tag{C1}$$

The associated filter coefficients are those of a second order resonant biquad filter,

$$\begin{cases} b_1 = -A_g, \\ b_2 = A_g, \\ a_1 = -2e^{-\pi B_g/F_s} \cos(2\pi F_g/F_s), \\ a_2 = e^{-2\pi B_g/F_s}, \end{cases} \tag{C2}$$

where $F_s$ is the sampling frequency and $F_g$, $B_g$, and $A_g$ are the centre frequency, bandwidth, and amplitude of the resonance (glottal formant) and are defined as

$$\begin{cases} F_g = \dfrac{1}{2O_qT_0}, \\ B_g = \dfrac{1}{O_qT_0 \tan(\pi(1-\alpha_m))}, \\ A_g = E. \end{cases} \tag{C3}$$

By setting $a_{\mathrm{CALM}} = \pi B_g$ and $b_{\mathrm{CALM}} = 2\pi F_g$, the time-domain impulse response of $\mathrm{LF}_{\mathrm{CALM}}$, truncated at $T_{CA} = -O_qT_0$, is given by computing the inverse $Z$-transform,

$$h_{\mathrm{CALM}open}(t) = -\frac{E}{\sin(\pi(1-\alpha_m))} e^{a_{\mathrm{CALM}}t} \times \sin\left(\frac{\pi}{O_qT_0} t + \pi(1-\alpha_m)\right), \quad t \in [-O_qT_0, 0]. \tag{C4}$$

The $\mathrm{LF}_{\mathrm{CALM}}$ closed phase causal filter is defined in in the $Z$-domain as

$$H_{ST}(z) = \frac{b_{ST}}{1 + a_{ST} z^{-1}}, \tag{C5}$$

and its filter coefficients are computed from the cut-off frequency $F_a = 1/(2\pi T_a)$,

$$\begin{cases} b_{ST} = 1 - e^{-2\pi F_a/F_s}, \\ a_{ST} = -e^{-2\pi F_a/F_s}. \end{cases} \tag{C6}$$

## APPENDIX D: DERIVATION OF $\mathrm{LF}_{\mathrm{LM}}$

$\mathrm{LF}_{\mathrm{LM}}$ is the causal version of $\mathrm{LF}_{\mathrm{CALM}}$ (Feugère *et al.*, 2017). Therefore, the glottal formant, also defined in the $Z$-domain, has the following transfer function:

$$H_{\mathrm{LM}_{open}}(z) = \frac{b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}}, \tag{D1}$$

whose coefficients are given by Eqs. (C2) and (C3). To have a convergent filter, it is necessary that $a_{\mathrm{LM}} < 0$. Therefore, $a_{\mathrm{LM}} = -\pi B_g$ and $b_{\mathrm{LM}} = 2\pi F_g$. Finally, the time-domain impulse response of $\mathrm{LF}_{\mathrm{LM}}$ is

$$h_{\mathrm{LM}_{open}}(t) = \frac{E}{\sin(\pi(1-\alpha_m))} e^{a_{\mathrm{LM}}t} \times \sin\left(\frac{\pi}{O_q T_0} t - \pi(1-\alpha_m)\right), \quad t > 0. \tag{D2}$$

The spectral tilt filter of $\mathrm{LF}_{\mathrm{LM}}$ is the same as $\mathrm{LF}_{\mathrm{CALM}}$ [Eqs. (C5) and (C6)].

## APPENDIX E: SYNTHESIS WITH $\mathrm{LF}_{\mathrm{CALM}}$ AND $\mathrm{LF}_{\mathrm{LM}}$

$\mathrm{LF}_{\mathrm{CALM}}$ open phase uses the anti-causal filter $H_{\mathrm{CALM}_{open}}$ [Eq. (C1)]. We define a pulse train $\delta_{gci}$ whose impulses are placed on the GCIs. The pulse train is then filtered by $H_{\mathrm{CALM}_{open}}$, leading to the recursion equation

$$x_{\mathrm{CALM}_{open}}[n] = b_1\delta_{gci}[n+1] + b_2\delta_{gci}[n+2] - a_1 x_{\mathrm{CALM}_{open}}[n+1] - a_2 x_{\mathrm{CALM}_{open}}[n+2]. \tag{E1}$$

For each period, the impulse response is truncated at the previous GOI. Then the full signal is filtered by the causal spectral tilt filter $H_{ST}$ [Eq. (C5)], leading to the recursion equation

$$x_{\mathrm{CALM}}[n] = b_{ST} x_{\mathrm{CALM}_{open}}[n-1] - a_{ST} x_{\mathrm{CALM}}[n-1]. \tag{E2}$$

In the case of $\mathrm{LF}_{\mathrm{LM}}$, both glottal formant and spectral tilt filters are applied in their causal form. We define a pulse train $\delta_{goi}$ whose impulses are placed on the GOIs. The pulse train is then filtered successively by the causal version of the glottal formant filter $H_{\mathrm{LM}_{open}}$ [Eq. (D1)], leading to the recursion equation

$$x_{\mathrm{LM}_{open}}[n] = b_1\delta_{goi}[n-1] + b_2\delta_{goi}[n-2] - a_1 x_{\mathrm{LM}_{open}}[n-1] - a_2 x_{\mathrm{LM}_{open}}[n-2], \tag{E3}$$

and the spectral tilt filter $H_{ST}$ [Eq. (C5)], leading to the recursion equation

$$x_{\mathrm{LM}}[n] = b_{ST} x_{\mathrm{LM}_{open}}[n-1] - a_{ST} x_{\mathrm{LM}}[n-1]. \tag{E4}$$

[1] http://cycling74.com (Last viewed 8/16/2021).
[2] See the supplementary material https://www.scitation.org/doi/suppl/10.1121/10.0005879 for all stimuli.


Airaksinen, M., Bollepalli, B., Juvela, L., Wu, Z., King, S., and Alku, P. (**2016**). "GlottDNN—A full-band glottal vocoder for statistical parametric speech synthesis," in *Proceedings of Interspeech*, September 8–12, San Francisco, CA, pp. 2473–2477.

Alku, P., Bäckström, T., and Vilkman, E. (**2002**). "Normalized amplitude quotient for parametrization of the glottal flow," J. Acoust. Soc. Am. **112**(2), 701–710.

Bozkurt, B., Doval, B., d'Alessandro, C., and Dutoit, T. (**2005**). "Zeros of Z-transform representation with application to source-filter separation in speech," IEEE Signal Process. Lett. **12**(4), 344–347.

Burkhardt, F., and Sendlmeier, W. F. (**2000**). "Verification of acoustical correlates of emotional speech using formant-synthesis," in *Proceedings of the ISCA Tutorial and Research Workshop on Speech and Emotion*, September 5–7, Newcastle, Northern Ireland, UK, pp. 151–156.

Cabral, J. P., Richmond, K., Yamagishi, J., and Renals, S. (**2014**). "Glottal spectral separation for speech synthesis," IEEE J. Sel. Top. Signal Process. **8**(2), 195–208.

Childers, D. G. (**1995**). "Glottal source modeling for voice conversion," Speech Commun. **16**(2), 127–138.

Childers, D. G., and Lee, C. K. (**1991**). "Vocal quality factors: Analysis, synthesis and perception," J. Acoust. Soc. Am. **90**(5), 2394–2410.

Crawley, M. J. (**2013**). *The R Book*, 2nd ed. (Wiley, New York), pp. 628–649.

d'Alessandro, N., d'Alessandro, C., Le Beux, S., and Doval, B. (**2006**). "Real-time CALM synthesizer: New approaches in hands-controlled voice synthesis," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, June 4–8, Paris, France, pp. 266–271.

Degottex, G., Lanchantin, P., Roebel, A., and Rodet, X. (**2013**). "Mixed source model and its adapted vocal tract filter estimate for voice transformation and synthesis," Speech Commun. **55**(2), 278–294.

Doval, B., d'Alessandro, C., and Henrich, N. (**2003**). "The voice source as a causal/anticausal linear filter," in *Proceedings of the ISCA Tutorial and Research Workshop on Voice Quality: Functions, Analysis and Synthesis*, August 27–29, Geneva, Switzerland, pp. 15–20.

Doval, B., d'Alessandro, C., and Henrich, N. (**2006**). "The spectrum of glottal flow models," Acta Acust. united Acust. **92**(6), 1026–1046.

Drugman, T., Bozkurt, B., and Dutoit, T. (**2011**). "Causal–anticausal decomposition of speech using complex cepstrum for glottal source estimation," Speech Commun. **53**(6), 855–866.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague, Netherlands), pp. 1–328.

Fant, G. (**1995**). "The LF-model revisited: Transformations and frequency domain analysis," Department for Speech, Music and Hearing Quarterly Progress and Status Report (KTH Computer Science and Communication, Stockholm, Sweden), Vol. 36, pp. 119–156.

Fant, G., Kruckenberg, A., Liljencrants, J., and Bavegard, M. (**1994**). "Voice source parameters in continuous speech: Transformation of LF-parameters," in *Proceedings of the International Conference on Spoken Language Processing*, September 18–22, Yokohama, Japan, pp. 1451–1454.

Fant, G., Liljencrants, J., and Lin, Q. (**1985**). "A four-parameter model of glottal flow," Department for Speech, Music and Hearing Quarterly Progress and Status Report 4 (KTH Computer Science and Communication, Stockholm, Sweden), Vol. 26, pp. 1–13.

Feugère, L., d'Alessandro, C., Doval, B., and Perrotin, O. (**2017**). "Cantor Digitalis: Chironomic parametric synthesis of singing," EURASIP J. Audio Speech Music Process. **2017**, 1.

Fujisaki, H., and Ljungqvist, M. (**1986**). "Proposal and evaluation of models for the glottal source waveform," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 7–11, Tokyo, Japan, Vol. 11, pp. 1605–1608.

Gardner, W. R., and Rao, B. D. (**1997**). "Noncausal all-pole modeling of voiced speech," IEEE Trans. Speech Audio Process. **5**(1), 1–10.

Gobl, C., Murphy, A., Yanushevskaya, I., and Ní Chasaide, A. (**2018**). "On the relationship between glottal pulse shape and its spectrum: Correlations of open quotient, pulse skew and peak flow with source harmonic amplitudes," in *Proceedings of Interspeech*, September 2–6, Hyderabad, India, pp. 222–226.

Gobl, C., and Ní Chasaide, A. (**2003**). "The role of voice quality in communicating emotion, mood and attitude," Speech Commun. **40**(1), 189–212.

Hedelin, P. (**1984**). "A glottal LPC-vocoder," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 19–21, San Diego, CA, pp. 21–24.

Henrich, N., d'Alessandro, C., and Doval, B. (**2001**). "Spectral correlates of voice open quotient and glottal flow asymmetry: Theory, limits and experimental data," in *Proceedings of Eurospeech*, September 3–7, Aalborg, Denmark, pp. 47–50.

Hézard, T., Hélie, T., and Doval, B. (**2013**). "A source-filter separation algorithm for voiced sounds based on an exact anticausal/causal pole decomposition for the class of periodic signals," in *Proceedings of Interspeech*, August 25–29, Lyon, France, pp. 54–58.

Holmes, J. (**1983**). "Formant synthesizers: Cascade or parallel?," Speech Commun. **2**(4), 251–273.
Juvela, L., Bollepalli, B., Tsiaras, V., and Alku, P. (**2019**). "GlotNet—A raw waveform model for the glottal excitation in statistical parametric speech synthesis," IEEE/ACM Trans. Audio Speech Lang. Process. **27**(6), 1019–1030.
Kingdom, F., and Prins, N. (**2016**). *Psychophysics: A Practical Introduction* (Academic, Cambridge, MA), pp. 1–346.
Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**(3), 971–995.
Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**(2), 820–857.
Markel, J. E., and Gray, A. H. (**1982**). *Linear Prediction of Speech* (Springer-Verlag, Berlin), pp. 1–290.
McLoughlin, I. V., Perrotin, O., Sharifzadeh, H., Allen, J., and Song, Y. (**2020**). "Automated assessment of glottal dysfunction through unified acoustic voice analysis," J. Voice, published online.
Ní Chasaide, A., Yanushevskaya, I., Kane, J., and Gobl, C. (**2013**). "The voice prominence hypothesis: The interplay of F0 and voice source features in accentuation," in *Proceedings of Interspeech*, August 25–29, Lyon, France, pp. 3527–3531.
Patel, S., Scherer, K. R., Björkner, E., and Sundberg, J. (**2011**). "Mapping emotions into acoustic space: The role of voice production," Biol. Psychol. **87**(1), 93–98.
Perrotin, O., and McLoughlin, I. (**2019**). "GFM-Voc: A real-time voice quality modification system," in *Proceedings of Interspeech*, September 15–19, Graz, Austria, pp. 3685–3686.
Perrotin, O., and McLoughlin, I. V. (**2020**). "Glottal flow synthesis for whisper-to-speech conversion," IEEE/ACM Trans. Audio Speech Lang. Process. **28**, 889–900.
Raitio, T., Suni, A., Vainio, M., and Alku, P. (**2013**). "Comparing glottal-flow-excited statistical parametric speech synthesis methods," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 26–31, Vancouver, Canada, pp. 7830–7834.
Rosenberg, A. E. (**1971**). "Effect of glottal pulse shape on the quality of natural vowels," J. Acoust. Soc. Am. **49**(2B), 538–590.
Veldhuis, R. (**1998**). "A computationally efficient alternative for the LF model and its perceptual evaluation," J. Acoust. Soc. Am. **103**(1), 566–571.