A simplified approximation of the four-parameter LF model of voice source

Yingyong Qi and Ning Bi
Department of Speech and Hearing Sciences, University of Arizona, Tucson, Arizona 85721

(Received 26 October 1993; accepted for publication 19 April 1994)

A simplified approximation of the four-parameter, voice source model developed by Fant *et al.* [STL-QPSR 4, 1–12 (1985)] is described in this Letter. In our approximation, the computational complexity required to implement the four-parameter model is simplified without reducing model accuracy. The simplified approximation accommodates rapid, on-line adjustment of source parameters during speech synthesis.

PACS numbers: 43.70.Gr, 43.72.Ja

INTRODUCTION

An accurate and versatile model of the human voice source is essential to the achievement of natural sounding synthetic speech (Fant *et al.*, 1985; Klatt and Klatt, 1990; Childers and Lee, 1991). Fant, Liljencrants, and Lin (1985) have developed a four-parameter model (hereafter referred to as the LF model) that has been widely used. The model has been used as an excitation input to a speech synthesizer and as a mathematical framework for analyzing different types of voices. The model is computationally complex. For example, it requires the solution of roots of two, nonlinear equations for each given set of model parameters. In addition, not all parameters in the model have direct physical correspondence with voice source events.

Our overall objective was to develop a simplified approximation of the LF model for use as excitation to a speech synthesizer. Specific goals were to reduce the computation needed to implement the LF model and to have all four parameters of the model correspond directly with physical events related to human voicing.

I. LF MODEL AND ITS APPROXIMATION

In the LF model, the flow derivative of the voice source is specified by four temporal parameters t_c, t_p, t_e, and t_a (see Fig. 1). Three of the parameters have direct physical correspondence with human voicing events: t_c could be the fundamental period, t_p is the instant of maximum flow, and t_e is the instance of maximum glottal closing. Parameter t_a is the second (right) derivative of the volume flow at the minimum of the first derivative. It does not have apparent physical correspondence with human voicing events. These parameters are referred to as the analysis set of parameters and they may be estimated from inverse filtering of the speech signal.

Another set of parameters, E_0, E_e, ω_g, α, and ϵ, are used to synthesize the flow derivative in the LF model. These parameters can be derived from the analysis set and appear in the model as

$$E(t)=\begin{cases} E_1(t)=E_0 e^{\alpha t}\sin\omega_g t \quad (t\leq t_e), \\ E_2(t)=(-E_e/\epsilon t_a)[e^{-\epsilon(t-t_e)}-e^{-\epsilon(t_c-t_e)}] \\ \qquad (t_e<t\leq t_c), \end{cases} \tag{1}$$

where $E(t)$ is the flow derivative.

All parameters in Eq. (1) need to be supplied explicitly in any numerical implementation of the LF model. Some parameters are relatively easy to obtain. For example, E_0 is an arbitrary gain constant and ω_g simply equals π/t_p. Others are not. For example, explicit values of α and ϵ are not available until roots of the following equations are solved:

$$1-e^{-\epsilon(t_c-t_e)}=\epsilon t_a, \tag{2}$$

$$\int_0^{t_c} E(t)=0. \tag{3}$$

Equation (2) results from setting $E_2(t_e)=-E_e$, where $E_e>0$ is the amplitude of flow derivative at t_e. Equation (3) ensures that the source flow function returns to baseline at the end of each period. Once α is obtained, E_e can be computed as

$$E_e=-E_0 e^{\alpha t_e}\sin(\omega_g t_e). \tag{4}$$

To avoid solving Eq. (3), Fant *et al.* approximate the area under the second segment of their model using $(E_e t_a/2)K_a$, where K_a is a function of t_c-t_e and t_a only. This substitution is referred to as the LF approximation. With this approximation, α becomes the root of an ordinary nonlinear equation,

$$\frac{e^{\alpha t_e}[\alpha\sin(\omega_g t_e)-\omega_g\cos(\omega_g t_e)]+\omega_g}{\alpha^2+\omega_g^2} \cong \frac{-e^{\alpha t_e}\sin(\omega_g t_e)t_a}{2}K_a, \tag{5}$$

resulting from flow continuity at t_e. Here, ϵ remains the root of Eq. (2). The roots of two, nonlinear equations for each given set of parameters also need to be solved to numerically compute source function by the LF approximation.

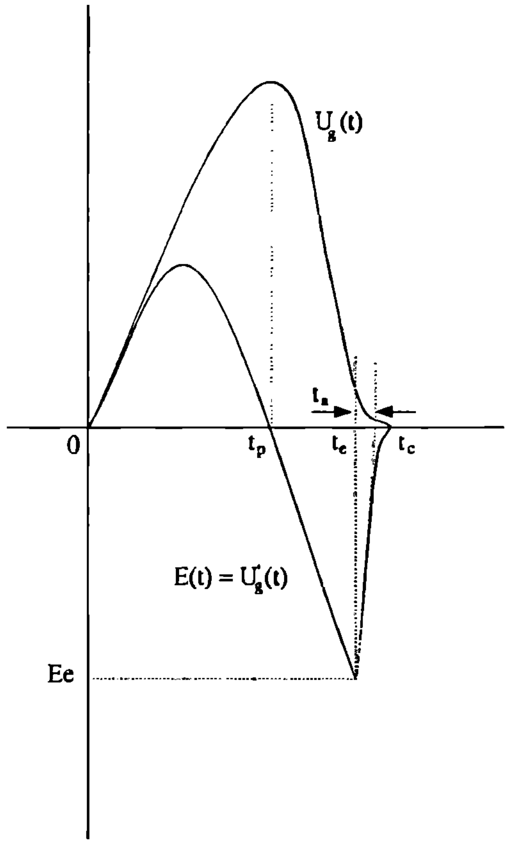

FIG. 1. The LF model.

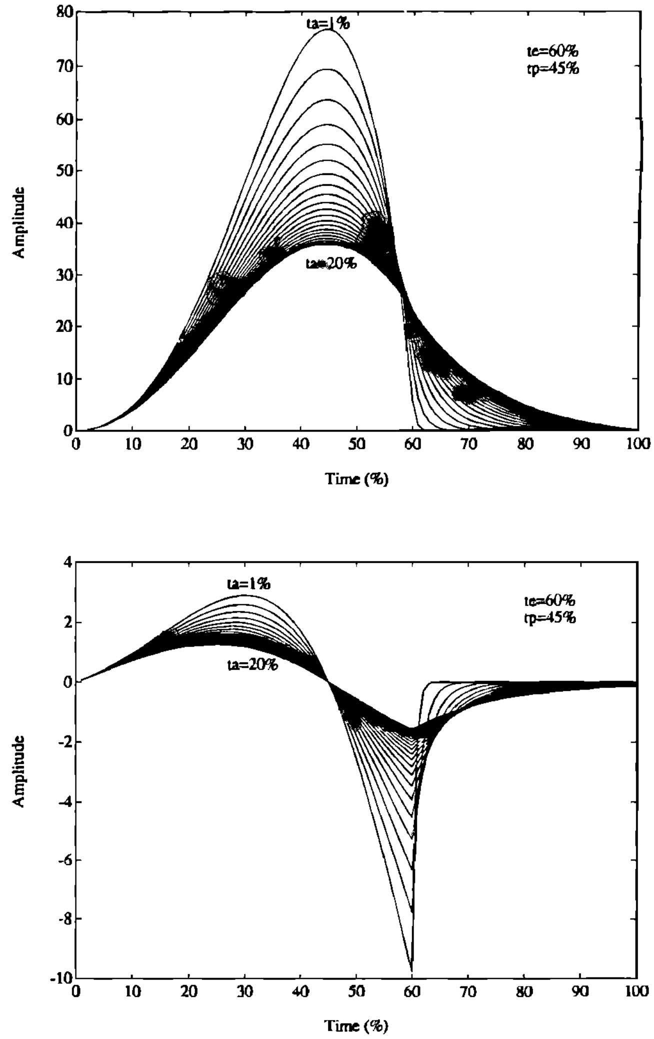

FIG. 2. (a) Flow of approximation I. (b) Flow derivative of approximation I.

II. A SIMPLIFIED APPROXIMATION

The complicated relationships among parameters of the LF model are largely related to the constraint imposed by Eq. (3). This constraint ensures that the flow returns to the baseline at the end of each period. This constraint also forces all parameter adjustments to be made concurrently on both segments of the LF model and imposes complicated relationships among model parameters.

In our simplified approximation, we eliminate t_a in the second segment of the LF model and use E_e as an explicit model parameter (Ananthapadmanabha and Fant, 1982; Ananthapadmanabha, 1984). As a result, the four parameters of the voice source model now include t_c, t_p, t_e, and E_e. In our approximation, the first segment of the LF model remains unchanged. In addition, the second segment of the LF model remains an exponential function, but takes a slightly different form:

$$E_2(t) = -E_e e^{-\epsilon(t-t_e)}, \tag{6}$$

where E_e is a known constant. The antiderivative of this equation is

$$U_2(t) = \int_{t_e}^{t} E_2(t) = \frac{E_e}{\epsilon} e^{-\epsilon(t-t_e)} - \frac{E_e}{e} e^{-\epsilon(t_c-t_e)}. \tag{7}$$

Using Eqs. (4) and (6), α is simply

$$\alpha = \frac{1}{t_e} \ln\left(\frac{-E_e}{E_0 \sin(\omega_g t_e)}\right) \tag{8}$$

based on Eq. (1). Here, ϵ is obtained by the flow continuity constraint,

$$U_1(t_e) = U_2(t_e), \tag{9}$$

where

$$U_1(t) = \int_0^t E_1(t). \tag{10}$$

Because $U_1(t)$ is not a function of ϵ [see the left-hand side of Eq. (5)], the root of only one, nonlinear equation needs to be solved to compute the source function. If it is further assumed that the closed phase is relatively long and the return to flow baseline is relatively fast following glottal closing [i.e., $\epsilon(t_c - t_e) \gg 1$], the term $e^{-\epsilon(t_c-t_e)}$ would be near zero. Under these assumptions, ϵ can be directly obtained as

$$\epsilon = E_e / U_1(t_e) \tag{11}$$

and no root solving is necessary.

III. EVALUATION OF THE APPROXIMATION

For purposes of comparison, the signal-to-noise (S/N) ratios of flow and flow derivative were computed, where the

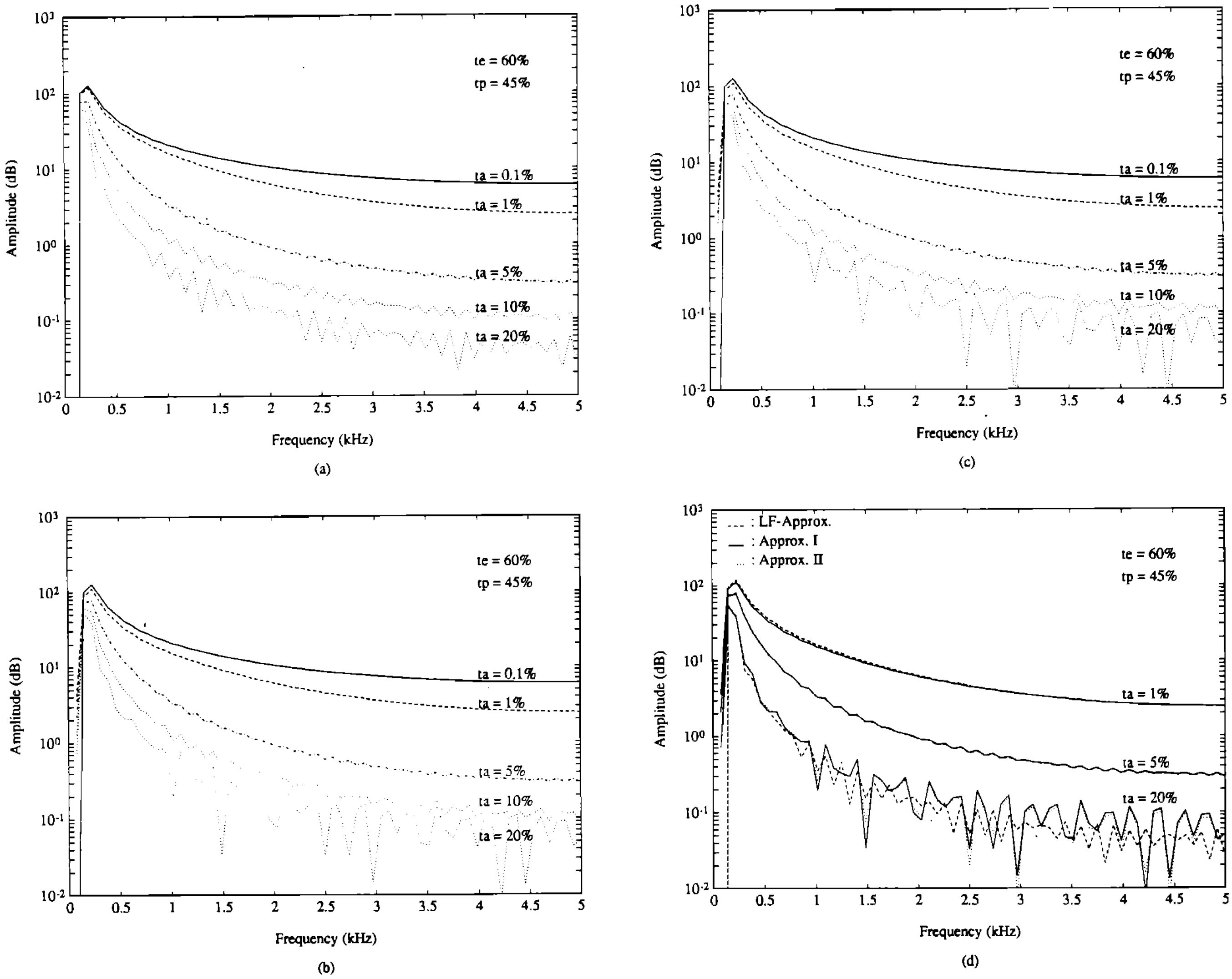

FIG. 3. Spectra of selected flow derivatives of (a) LF model, (b) approximation I, (c) approximation II, and (d) LF model and its approximations.

S/N ratio is defined as the ratio between the rms energy of the original LF model and the rms energy of the amplitude difference between the LF model and our approximation. The S/N ratios were computed for the approximations described in Sec. II.

In these evaluations, α and ϵ were obtained using Eqs. (2) and (3) for the LF model. The integration required in Eq. (3) was accomplished using forward Euler's method. For our simplified approximation, α was obtained using Eq. (8). Here, ϵ was obtained using Eq. (9) or (11) depending on the desired level of approximation. When Eq. (9) was used, we refer to the result as approximation I. When Eq. (11) was used, we refer to the result as approximation II. A bi-section method was used to solve all necessary roots in the evaluation process (Dahlquist *et al.*, 1974). For easy comparison, the gain constant, E_0, was set to unity and t_a was always used to denote the evaluation results in place of E_e, because t_a and E_e are equivalent independent parameters for our simplified approximations.

The flow and flow derivative at various parameter values for approximation I are illustrated in Fig. 2. Here, t_c, t_e, and t_p were held constant, and t_a was varied from 1% to 20%[1] of the fundamental period (t_c). Plots for the LF model and approximation II are not illustrated because they were nearly identical to the plots for approximation I. Spectra of selected flow derivatives are shown in Fig. 3(a), (b), and (c) for the LF model, approximation I, and approximation II, respectively. For convenience of comparison, these spectra are merged in Fig. 3(d). These results demonstrate that our simplified approximations are capable of producing a wide range of source functions characterized by temporal and spectral content that is comparable to that produced by the LF model.

Some S/N ratios of the approximated flow and flow derivative are shown in Fig. 4 for various parameter models. Here, t_a was 1%–20% and t_p was 40%–50% of the fundamental period. Additional S/N ratios of the approximated flow and flow derivative are shown in Fig. 5. Here, t_a was 1%–20%, while t_e was 55%–65% of the fundamental period.

Evaluation results summarized in Figs. 2–5 show that the differences between approximation I and II are relatively small when t_a is small. Observable differences exist between approximation I and II when t_a is relatively large (see Figs. 4 and 5). The differences between approximation I and II at

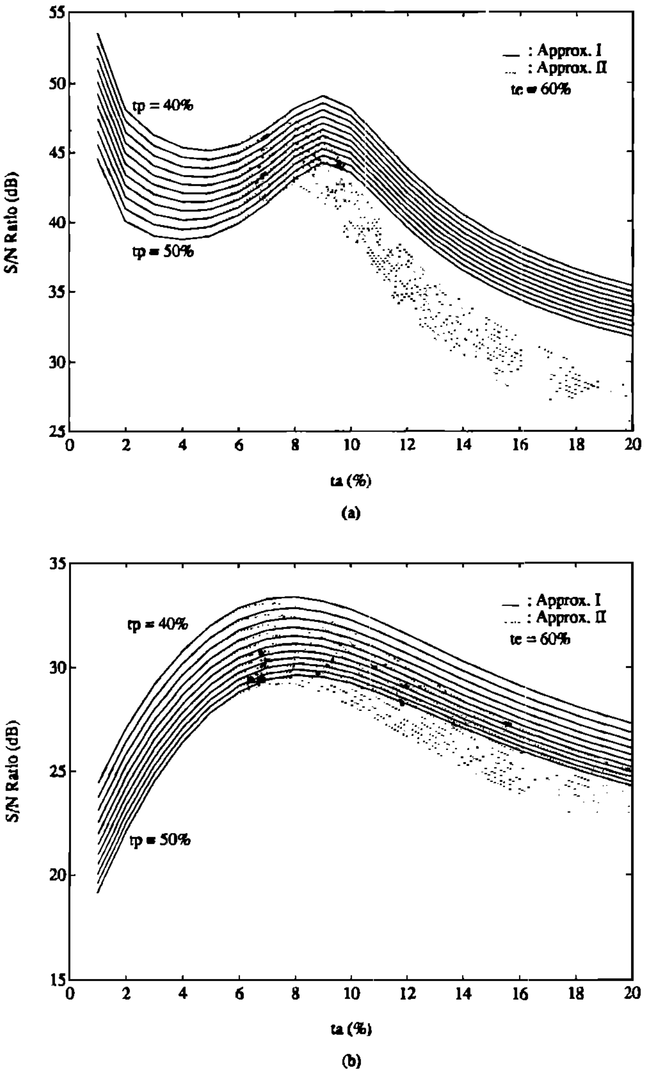

FIG. 4. (a) Signal-to-noise ratios of flow approximations. (b) Signal-to-noise ratios of flow derivative approximations.

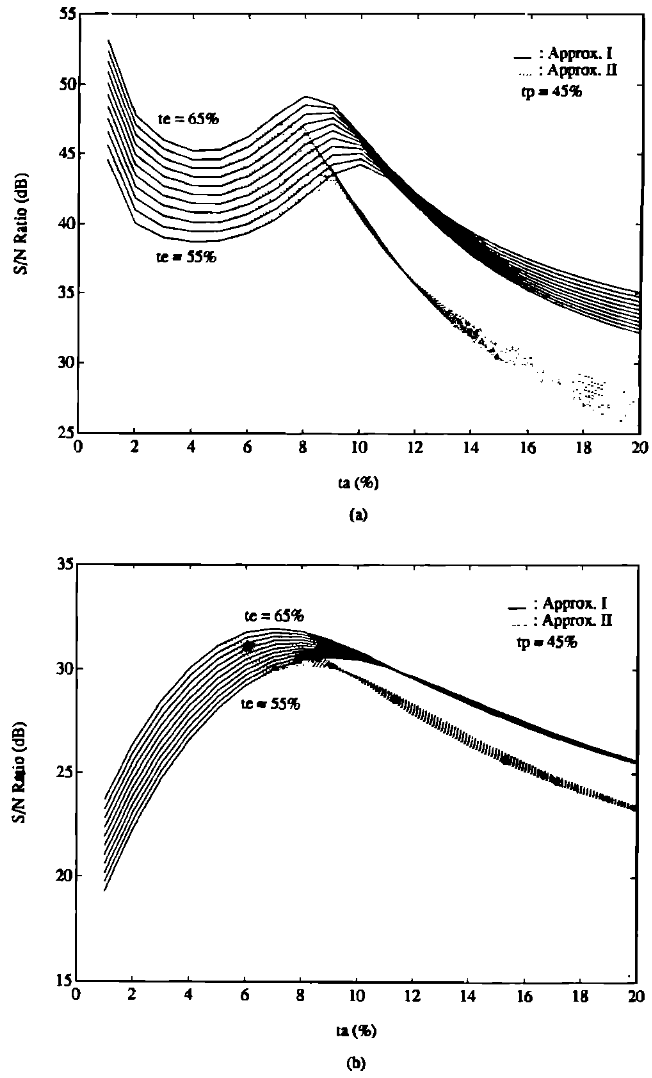

FIG. 5. Signal-to-noise ratios of flow approximations. (b) Signal-to-noise ratios of flow derivative approximations.

large t_a values are due to the need for the flow function to return to the baseline at t_c in approximation I [see Eq. (7)], while the flow function approaches the baseline asymptotically in approximation II. These evaluation results also indicate that the S/N ratios are generally higher for flow than for flow derivative. The approximation of the flow is most accurate when t_a is relatively small (<10% of the pitch period). The approximations of flow derivative are most accurate when t_a is in the middle range and not in an extreme condition.

Approximations I and II are based on flow rather than flow derivative. Use of the integrated forms of these approximations is recommended whenever possible. For example, during speech synthesis, the derivation of the synthetic signal should be made in the final operation of the synthesizer. On the basis of informal listening, we could not detect any differences between speech synthesized using approximations I and II and speech synthesized using the LF model. The simplifications of approximations I and II permit rapid, on-line adjustment of source parameters during speech synthesis.

ACKNOWLEDGMENT

This work was supported, in part, by a grant from the National Institute of Deafness and Other Communication Disorders, DC01440, Analysis and Improvement of Alaryngeal Speech. We appreciate the assistance of Dr. Bernd Weinberg in preparing this Letter.

[1]Note that the corresponding E_e value of each t_a is available from the flow derivative plots.

Ananthapadmanabha, T. (1984). "Acoustic analysis of voice source dynamics," STL-QPSR 2–3, 1–24.

Ananthapadmanabha, T., and Fant, G. (1982). "Calculation of true glottal flow and its components," STL-QPSR 1, 1–30.

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

Dahlquist, G., Bjorck, A., and Anderson, N. (1974). *Numerical Methods* (Prentice-Hall, Englewood Cliffs, NJ).

Fant, G., Liljencrants, J., and Lin, Q. G. (1985). "A four-parameter model of glottal flow," STL-QPSR 4, 1–12.

Klatt, D., and Klatt, L. (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. 87, 820–857.