A computationally efficient alternative for the Liljencrants–Fant model and its perceptual evaluation

Raymond Veldhuis[a)]
IPO—Centre for Research on User-System Interaction, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

(Received 4 June 1997; accepted for publication 17 September 1997)

An alternative for the Liljencrants–Fant (LF) glottal-pulse model is presented. This alternative is derived from the Rosenberg model. Therefore, it is called the Rosenberg++ model. In the derivation a general framework is used for glottal-pulse models. The Rosenberg++ model is described by the same set of T or R parameters as the LF model but it has the advantage over the LF model that it is computationally more efficient. It is compared with the LF model in a psychoacoustic experiment, from which it is concluded that in a practical situation it is capable of producing synthetic speech which is perceptually equivalent to speech generated with the LF model.
 [S0001-4966(98)00701-2]

PACS numbers: 43.70.Gr, 43.72.Ja, 43.71.Bp [AL]

INTRODUCTION

For analysis and synthesis purposes, speech production is often modeled by a source-filter model. Figure 1 shows two versions of such a source-filter model. On the left we see a model consisting of a source producing a signal $g(t)$, which models the airflow passing the vocal cords, a filter with a transfer function $H(j\omega)$, which models the spectral shaping by the vocal tract, and an operator R, which models the conversion of the airflow to a pressure wave $s(t)$ at the lips and which is called lip radiation. The operator R is essentially a differentiation operator. On the right we see an equivalent model, in which the differentiation operator has been combined with the source, which now produces the time derivative $\dot{g}(t)$ of the airflow passing the vocal cords. The opening between the vocal cords is called the glottis, therefore the source is referred to as the glottal source. In voiced speech the signal $g(t)$ is periodic and one period of $g(t)$ is called a glottal pulse. The glottal pulse or, more often, its time derivative has been the topic of many studies because it is expected to determine the voice quality and to be related to the production of prosody (e.g., Childers and Lee, 1991; Cummings and Clements, 1993; Gobl, 1989; Klatt and Klatt, 1990; Pierrehumbert, 1989; Rosenberg, 1971; Strik, 1994). The time derivative of the glottal pulse is studied rather than the glottal pulse because it is more easily obtained from the speech signal and some of the glottal-source parameters can be more easily derived from it.

The Liljencrants–Fant (LF) model (Fant *et al.*, 1985) has become a reference model for glottal-pulse analysis. Unfortunately, its use in speech synthesizers is limited because of its computational complexity. This computational complexity is mostly due to the difference between the specification parameters and the generation parameters of the LF model. The computation of the generation parameters from the specification parameters is computationally complex, because it involves solving a nonlinear equation. This is explained in Sec. I. This section presents a general framework for glottal-pulse models and also introduces the LF model. In Sec. II we introduce the Rosenberg++ (R++) model, which has the same features as the LF model but has the advantage that it can be computed directly and simply from the specification parameters. Section III describes an experiment which shows to what extent the R++ model is capable of generating synthetic speech that is perceptually equivalent to speech generated with the LF model. The computational complexities of the R++ and the LF model are compared in Sec. IV. Finally, Sec. V presents conclusions.

I. GLOTTAL-PULSE MODELS

Figure 2 shows typical examples of the glottal waveform $g(t)$ and its time derivative $\dot{g}(t)$ and it introduces the specification parameters t_0, t_p, t_e, t_a and U_o or E_e (Fant *et al.*, 1985). The length of a glottal cycle is t_0. The maximum airflow U_o occurs at t_p and the maximum excitation with amplitude E_e occurs at t_e, which corresponds to the instant when the vocal cords collide. The interval before t_e is called the open phase. The interval with approximate length $t_a = E_e/\ddot{g}(t_e)$ just after the instant of maximum excitation is called the return phase. The interval between t_e, or better $t_e + t_a$, and the end of the glottal cycle is called the closed phase. During this phase the vocal folds have reached maximum closure and the airflow has reduced to its minimum. A minimum airflow greater than 0 is often referred to as leakage (Holmberg *et al.*, 1988). The presence of leakage influences the speech in two ways. First, by introducing a permanent acoustic coupling between vocal tract and trachea it influences the formant resonances and thus the speech spec-

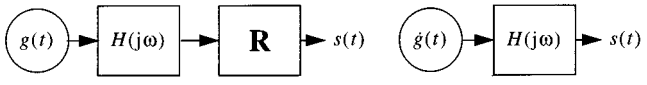

FIG. 1. Left: Source-filter model with glottal source, vocal-tract filter, and lip radiation. Right: Equivalent source-filter model with lip radiation and glottal source combined.

a)Electronic mail:veldhuis@ipo.tue.nl

trum. Second, it may affect the glottal excitation by introducing turbulence noise. Due to the differentiating effect of the lip radiation, the constant additional airflow has no effect on the pressure wave at the lips. In a speech synthesizer based on a simple source-filter model, the effects of leakage can be accounted for by adding a noise term to the source and by adapting the transfer function of the filter. Therefore, we assume that there is no leakage and that $g(0)=g(t_0)=0$. The airflow in the return phase is generally considered to be of perceptual importance, because it determines the spectral slope. The parameters t_0, t_p, t_e, and t_a are called the T parameters. Instead of the T parameters, sometimes the R parameters (Fant *et al.*, 1985) are used, which are defined as follows:

$$r_o=t_e/t_0, \quad r_a=t_a/t_0, \quad r_k=(t_e-t_p)/t_p. \tag{1}$$

The parameters r_o and r_a denote the relative duration of the open phase and the return phase, respectively. The parameter r_k quantifies the symmetry of the glottal pulse.

The following expression is a general description of the glottal-pulse time derivative $\dot{g}(t)$ with an exponential decay modeling the return phase:

$$\dot{g}(t)=\begin{cases} f(t), & \text{for } 0\leqslant t<t_e, \\ f(t_e)\,\dfrac{\exp(-(t-t_e)/t_a)-\exp(-(t_0-t_e)/t_a)}{1-\exp(-(t_0-t_e)/t_a)} & \\ & \text{for } t_e\leqslant t<t_0. \end{cases} \tag{2}$$

The function $f(t)$ on the first line of (2) describes the glottal-pulse time derivative until the instant of excitation. The second line models the return phase. Integration leads to the following expression for the glottal airflow:

$$g(t)=\begin{cases} \displaystyle\int_0^t f(\tau)\,d\tau, & \text{for } 0\leqslant t<t_e, \\ \displaystyle\int_0^{t_e} f(\tau)\,d\tau+t_a f(t_e)\frac{1-\exp(-(t-t_e)/t_a)-[(t-t_e)/t_a]\exp(-(t_0-t_e)/t_a)}{1-\exp(-(t_0-t_e)/t_a)} & \text{for } t_e\leqslant t<t_0. \end{cases} \tag{3}$$

Since there is no leakage we require $g(t)\geqslant 0$ and $g(0)=g(t_0)=0$, from which one can derive the following continuity condition:

$$\int_0^{t_e} f(\tau)\,d\tau+t_a f(t_e)D(t_0,t_e,t_a)=0, \tag{4}$$

with

$$D(t_0,t_e,t_a)=1-\frac{(t_0-t_e)/t_a}{\exp((t_0-t_e)/t_a)-1}. \tag{5}$$

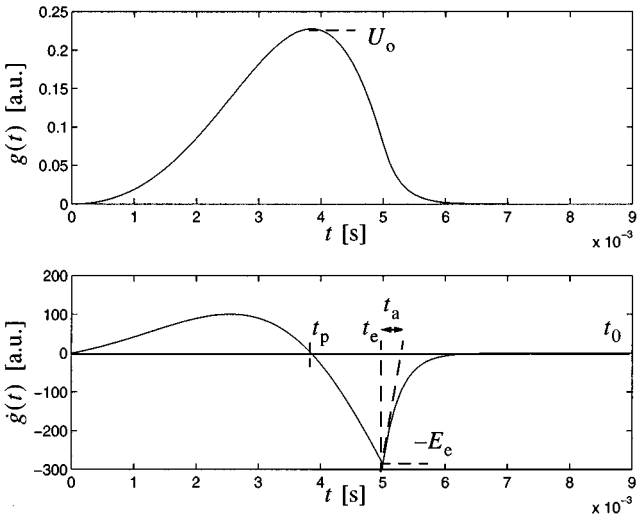

FIG. 2. Glottal pulse (top) and its time derivative (bottom). Arbitrary units.

Any parameters of $f(t)$ must be chosen such that condition (4) is satisfied.

The parameter t_a in the above definitions for the glottal airflow $g(t)$ and its time derivative $\dot{g}(t)$ is the time constant of the exponential decay in the return phase. This is slightly different from the situation in Fig. 2 and the original definition of the LF model (Fant *et al.*, 1985), in which t_a specifies the initial slope of the return phase. In Fant *et al.* (1985) the factor $1/t_a$ in the exponential functions is replaced by another generation parameter ε, determining the time constant of the decay. For $t_a\ll t_0-t_e$, which is usually the case, we have $\varepsilon=1/t_a$. Otherwise there exists a simple relation between t_a and ε (Fant *et al.*, 1985), which makes the two definitions equivalent.

The above set of expressions (2)–(5) forms a general framework for glottal-pulse models with an exponential return phase. The LF model follows from this framework for a specific choice of $f(t)$ and so does the Rosenberg model. Other models can be constructed by choosing other functions $f(t)$. Even other choices for the return phase are possible, but will not be discussed here.

The LF model presented in Fant *et al.* (1985), but with the modified definition of t_a, follows from (2) and the choice

$$f(t)=B\,\sin\left(\pi\,\frac{t}{t_p}\right)\exp(\alpha t), \tag{6}$$

with B the amplitude of the glottal-pulse time derivative. The generation parameter α can only be solved numerically from the continuity equation (4), which in this case reads

$$\frac{\pi-\exp(\alpha t_e)(\pi\cos(\pi(t_e/t_p))-\alpha t_p\sin(\pi(t_e/t_p)))}{\pi^2+(\alpha t_p)^2}$$

$$+\frac{t_a}{t_p}\exp(\alpha t_e)\sin\left(\pi\frac{t_e}{t_p}\right)D(t_0,t_e,t_a)=0. \quad (7)$$

Solving (7) for α is a heavy computational load in a speech synthesizer, where the T parameters may vary typically every 10 ms.

Qi and Bi (1994) propose an approximation in which t_0, t_p, t_e, and E_e are specified rather than t_0, t_p, t_e, and t_a. In this manner (7) needs not to be solved, but it has the disadvantage that the control over the return phase is lost. The R++ model, which is introduced in the next section, allows specification of t_0, t_p, t_e, and t_a and it does not require the solution of (7). One can ask whether the control over the return phase can be maintained without having to solve (7) with the other LF parameters, e.g., $F_a=1/(2\pi t_a)$, $F_g=1/(2t_p)$, and $r_g=t_0/(2t_p)$ (Fant *et al.*, 1985; Fant, 1995). Unfortunately, these parameters have simple relations to the T or the R parameters and, therefore, their use still requires the solution of an equation very similar to (7).

II. THE ROSENBERG++ MODEL

The R++ model is an extension of the Rosenberg model (Rosenberg, 1971):

$$g(t)=\begin{cases}At^2(t_e-t) & \text{for } 0\leqslant t<t_e,\\ 0 & \text{for } t_e\leqslant t<t_0,\end{cases}$$
$$\dot{g}(t)=\begin{cases}3At(\frac{2}{3}t_e-t) & \text{for } 0\leqslant t<t_e,\\ 0 & \text{for } t_e\leqslant t<t_0,\end{cases} \quad (8)$$

with A the amplitude of the glottal pulse. The Rosenberg model does not have a return phase and always has $t_p=2t_e/3$, or $r_k=1/2$. This limits its flexibility. Sometimes a pseudo return phase is introduced by applying a first- or second-order recursive low-pass filter to the glottal-pulse time derivative (e.g., Klatt and Klatt, 1990), but this undesirably changes t_p. We propose two extensions to the Rosenberg model, each explaining a ''+'' in ''R++.'' The first is a return phase as in (2). The second is an extra factor in $f(t)$, which allows us to specify a t_p. The latter extension results in

$$f(t)=4At(t_p-t)(t_x-t),$$
$$\int_0^t f(\tau)\,d\tau=At^2\left(t^2-\frac{4}{3}t(t_p+t_x)+2t_pt_x\right). \quad (9)$$

The following expression for t_x follows on solving the continuity equation (4):

$$t_x=t_e\left(1-\frac{\frac{1}{2}t_e^2-t_et_p}{2t_e^2-3t_et_p+6t_a(t_e-t_p)D(t_0,t_e,t_a)}\right). \quad (10)$$

The denominator of (10) vanishes when

$$t_p=\frac{2}{3}t_e\frac{t_e+3t_aD(t_0,t_e,t_a)}{t_e+2t_aD(t_0,t_e,t_a)}. \quad (11)$$

In that case, the R++ model reduces to an R+ model:

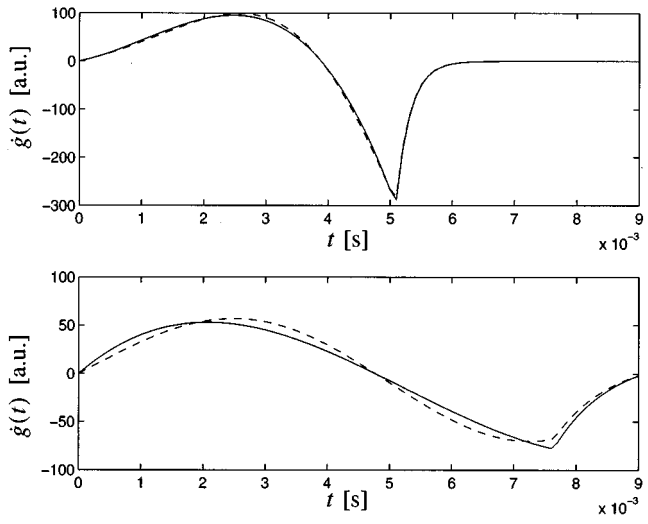

FIG. 3. The LF (dashed lines) and R++ (solid lines) glottal-pulse time derivatives. Top panel: $f_0=1/t_0=110$ Hz, $r_0\neq0.56$, $r_k=0.31$, and $r_a=0.025$. Bottom panel: $f_0=110$ Hz, $r_0=0.84$, $r_k=0.60$, and $r_a=0.10$. Arbitrary units.

$$f(t)=3At(t_p-t), \quad \int_0^t f(\tau)\,d\tau=At^2\left(\frac{3}{2}t_p-t\right), \quad (12)$$

which is the Rosenberg model extended with a return phase. The influence of the factor t_x-t on the shape of the glottal-pulse time derivative decreases with increasing $|t_x|$. For $|t_x|$ large enough, therefore, the R++ model can be replaced by the R+ model (12). The following condition guarantees that either $t_x\geqslant t_e$ or $t_x\leqslant0$:

$$\frac{1}{2}t_e\leqslant t_p\leqslant\frac{3}{4}t_e\left(\frac{t_e+4t_aD(t_0,t_e,t_a)}{t_e+3t_aD(t_0,t_e,t_a)}\right). \quad (13)$$

It ensures that $g(t)$ is non-negative. The left-hand inequality of (13) gives a lower limit to the skewness of the glottal pulse, which makes it symmetrical when we ignore the return phase. This limit is also known in the LF model. The right-hand inequality of (13) is an upper limit to the glottal-pulse skewness which is not present in the LF model, but so far has not turned out to be a limitation.

Figure 3 shows LF (dashed lines) and R++ (solid lines) glottal-pulse time derivatives for two sets of R parameters. The top panel shows glottal-pulse time derivatives for a modal voice with a distinct closed phase and the bottom panel for an abducted voice without a distinct closed phase. All waveforms have been power normalized to the same value. In general, the R++ waveform approximates the LF waveform closely if $r_k<0.5$. For higher r_k, the approximation is slightly worse. In any case, the differences between the models are small compared with the differences between the LF model and estimated waveforms (e.g., Childers and Lee, 1991; Strik, 1994). Therefore, we can already assume that both models are equally useful. However, because we want to present the R++ model as a computationally efficient alternative to the LF model with the same specification parameters, we believe it relevant that both models are also perceptually equivalent. This is investigated in Sec. III. The computational complexity is then addressed in Sec. IV.

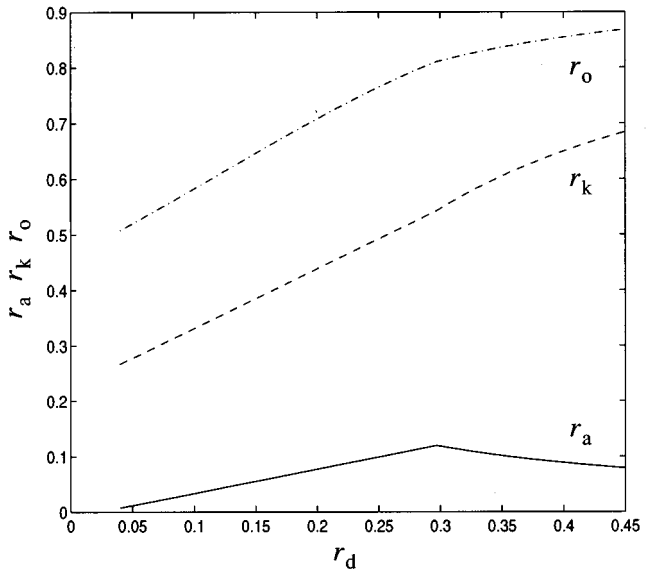

FIG. 4. Statistical dependencies of r_a, r_k, and r_o on r_d.

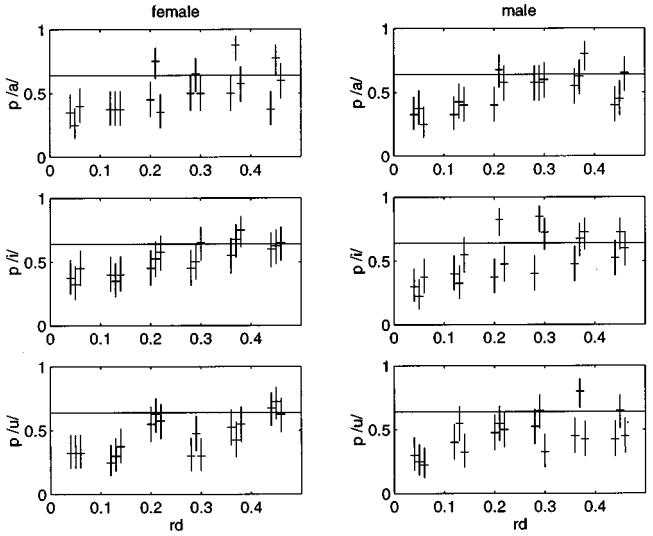

FIG. 5. Fractions correct and 90% confidence intervals, presented in triples for all subjects. Left: EK, middle; RK, right: SP. The values of subject RK are positioned at the correct r_d values. The horizontal line at $p=0.64$ corresponds to $d'=1$.

III. A PERCEPTUAL COMPARISON

By means of a psycho-acoustic experiment we investigated whether synthetic vowels generated with the R++ and the LF model at various choices of the *R* parameters can be perceptually discriminated. We chose this approach because a psycho-acoustical comparison of isolated vowels is more critical with respect to discrimination than the comparison of synthetic speech in which other synthesis artifacts and the context may mask the perceptual differences.

In order to choose *R* parameters corresponding to those of natural voices, we used the so-called shape parameter

$$r_d=\frac{U_0}{E_e t_0}, \tag{14}$$

which has been defined in Fant *et al.* (1994) and has been discussed extensively in Fant (1995). The definition above differs from the one in Fant *et al.* (1994) by a normalization factor. In Fant *et al.* (1994) and Fant (1995) it is observed that there exist simple statistical relations between r_d and the other *R* parameters, such that each of the *R* parameters can be predicted from a measured value of r_d. These relations are shown in Fig. 4. We chose the set {0.05, 0.13, 0.21, 0.29, 0.37, 0.45} as the values for r_d and used Fig. 4 to determine the *R* parameters.

From recordings of one male and one female voice we derived formant filters and fundamental frequencies $f_0=1/t_0$ for the vowels /a/, /i/, and /u/. Segments of 0.3 s of these vowels were synthesized for the six values of r_d with the simplified source-filter model of Fig. 1. The glottal-pulse time derivatives were according to the LF and the R++ models, respectively. The fundamental frequencies and formant filters were kept identical to those obtained from the recordings. The fundamental frequencies of the male and female vowels were approximately 110 and 200 Hz, respectively. The sampling frequency was 8 kHz. This resulted in 36 pairs of stimuli.

Three subjects (EK, RK, and SP), two of whom (RK and SP) had experience with psycho-acoustic experiments, were asked to discriminate between the segments generated with the LF and the R++ models in a three-interval three-alternative forced-choice paradigm. Each triple of stimuli was presented 40 times in a random order. In half of the trials the two reference segments were generated with the LF model and the subject's task was to indicate the segment generated with the R++ model. In the other half of the trials the two reference segments were generated with the R++ model and the subject's task was to indicate the segment generated with the LF model. The numbers of correct answers were registered. The experiment was split up into two sessions of equally many trials. The subjects were seated in a sound-proof booth. The stimuli were presented over Beyer DT990 headphones at an overall level of 70 dB SPL with a roving of 1 dB. The subjects received immediate feedback after every trial.

There was no significant difference between the results of the trials with the LF model and those with the R++ model in the reference intervals. Only subject RK showed a small learning effect in the second session, which has been ignored in the presentation of the results. Figure 5 shows a graphical presentation of the fractions correct p with 90% confidence intervals for all the subjects and all the pairs of stimuli as a function of r_d. The fractions correct are plotted in triples, the left, middle, and right one corresponding to subjects EK, RK, and SP, respectively. The values of subject RK are positioned at the correct r_d values. The horizontal line at $p=0.64$ corresponds to $d'=1$. This is the value at which the subjects are assumed to be able to just detect the differences (Green and Swets, 1966).

The results in Fig. 5 show that subject RK sometimes, and the other subjects only very occasionally, detected a difference. These detections occur at the higher values of r_d, where, because of the higher r_k, the waveform approxima-

tion is less good. In general, the fractions correct seem to increase somewhat with increasing r_d.

An interpretation of the 90% confidence interval is the following. If we assume that a subject has a detection probability p, then a hypothesis test based on the data would accept all values of p in the 90% confidence interval with a Type I error of 5%. This implies that when a confidence interval is entirely below the line $d'=1$, the data confirms that there is no discrimination. When a confidence interval is entirely above the line $d'=1$, the data confirms that there is certain discrimination. Finally, when a confidence interval crosses the line $d'=1$, discrimination is undetermined. In this view, certain discrimination is rare. In most cases, especially for lower r_d values below 0.2, there is no discrimination and for the higher r_d values discrimination is either absent or undetermined. On the basis of these observations and the fact that a psycho-acoustical experiment is more critical with respect to discrimination than a comparison of synthetic speech, we believe that, although the results show that the models are not completely perceptually equivalent, it is unlikely that there will be audible differences between synthetic speech generated with the R++ and with the LF model.

IV. COMPUTATIONAL EFFICIENCY

A general comparison of the computational efficiency of the LF and the R++ model is difficult, because its result depends on the specific implementations and on the speech-synthesis hardware. Therefore, we will focus on one aspect of the computational efficiency which is more or less hardware independent: the computational load in terms of numbers of the basic operations addition (+/−), multiplication (×), division (/) and function evaluation (sin, cos, exp). In addition, processing times of implementations of the two models in C on a RISC-4000 processor with a floating-point unit will be presented as an example.

The computational load of a glottal-pulse model breaks down into the computational load of calculating the samples of the glottal-pulse derivative and that of updating the parameters α for the LF model and t_x for the R++ model. Fant *et al.* (1985) and Lin (1990) propose the second-order recursive expression

$$s_n = 2e^{\alpha T_s}\cos\left(\pi\frac{T_s}{t_p}\right)s_{n-1} - (e^{\alpha T_s})^2 s_{n-2} = a_1 s_{n-1} + a_2 s_{n-2}. \tag{15}$$

for the calculation of the samples before t_e and the first-order recursive expression

$$s_m = e^{-T_s/t_a}s_{m-1} - \frac{(1-e^{-T_s/t_a})e^{-(t_0-t_e)/t_a}}{1-e^{-(t_0-t_e)/t_a}} = \rho s_{m-1} + c \tag{16}$$

for the samples after t_e. Here T_s is the sampling period. The evaluation of (15) requires two multiplications and one addition. The initial values s_0 and s_1 have to be reset at the beginning of a glottal cycle. The evaluation of (16) requires one multiplication and one addition. The R++ glottal-pulse derivative before t_e, written as $t(t-t_p)(t-t_x)$, requires two multiplications and three additions per sample. The number of additions is three because each factor is incremented by T_s. The R++ glottal-pulse derivative after t_e is computed recursively by using (16).

TABLE I. Computational load of glottal-pulse models in numbers of basic operations +, ×, /, function evaluation $f(.)$, and measured average processing times. The number of iterations to compute α is denoted by N_{it}.

	+	×	/	$f(.)$	Processing times (μs)
LF sample	1	r_0+1	0	0	9.4
R++ sample	$2r_0+1$	r_0+1	0	0	13.6
LF update	$5+4N_{\text{it}}$	$10+4N_{\text{it}}$	7	$8+N_{\text{it}}$	57.8
R++ update	8	8	3	1	10.7

In an LF update, $D(t_0,t_e,t_a)$ defined in (5) is computed and α is solved from (7). When α is known, the coefficients and the initial values of (15) and (16) are computed. The computation of $D(t_0,t_e,t_a)$ requires three additions, two divisions, and one function evaluation. Equation (7) is rewritten as

$$Q(u) = 1 + e^{\varphi u}\left(\left(\frac{t_a}{t_e}\,\varphi D(t_0,t_e,t_a)(1+u^2)+u\right)\times\sin(\varphi)-\cos(\varphi)\right), \tag{17}$$

with $\varphi=\pi t_e/t_p$, and solved for u. The parameter α is then found as $\alpha=u\pi/t_p$. The computation of φ, the parts of (17) independent on u, and α from u require four multiplications, two divisions, and two function evaluations. We propose the following iterative procedure for the finding the zero of (17). The function $Q(u)$ decreases with increasing u. First, if $Q(0)>0$, starting from $u=1$, u is doubled until $Q(u)<0$, or, if $Q(0)\leqslant 0$, starting from $u=-1$, u is doubled until $Q(u)>0$. Second, the interval in which the sign change occurs is searched for the zero by a binary-search algorithm. This requires a total number of N_{it} iterations, depending on the specification parameters and the required accuracy. Each iteration requires four additions, four multiplications, and one function evaluation. Computing coefficients and the initial values of (15) and (16) requires two additions, six multiplications, three divisions, and five function evaluations. In an R++ update $D(t_0,\ t_e,\ t_a)$ is computed and then t_x, which requires five additions, eight multiplications, and one division.

The results are summarized in the first four columns of Table I. In a formant synthesizer the computational loads of the LF and the R++ model for computing the samples are both small compared with that of the formant filters. With respect to computing the samples, the LF model is somewhat more efficient than the R++ model because of the smaller number of additions. The advantage of the R++ model is that the computational load for the parameter updates is smaller for any value of $N_{\text{it}}\geqslant 1$ and much smaller for realistic values of N_{it}, which are around 16 for a maximum relative error in α of 10^{-4}.

As an example, we measured the processing times of C-language implementations for the computation of the updates and one glottal cycle on a 133-MHz RISC-4000 processor with a floating-point unit. In a speech synthesizer run-

ning on a desk-top computer this determines the overall computational complexity, because the glottal cycle can be stored and played out repeatedly until the next parameter update. The sampling frequency was 10 000 Hz, the length of a glottal cycle was 9 ms, and the maximum relative error in α was 10^{-4}. The results were calculated for the r_d values of Sec. IV. The averages are presented in the last column of Table I. The processing times for the samples increased by about 4% for the LF model and 15% for the R++ model with increasing r_d, and, consequently, r_o. These increases and their differences were to be expected on account of the r_o dependence shown in Table I. The average processing time for the samples of the LF model is about 30% less than that of the R++ model, which is due to the larger numbers of additions before t_e of the R++ model. Considering the average processing times for the updates, we see that the R++ model is 5.4 times faster than the LF model. In this example the number of iterations N_{it} was equal to 16 for all values of r_d. The R++ model was overall about 2.8 times faster than the LF model. This number will be higher for voices with shorter glottal cycles.

In addition to the smaller overall computational load, the R++ model has two other advantages over the LF model, which are relevant for an implementation in dedicated hardware. The first is that the computation of t_x is data independent and straightforward, whereas the computation of α is data dependent and involves an iterative search for a zero, requiring additional decision and control logic. The second additional advantage of the R++ model is that, because of the shorter time spent on updates, the samples can be put out more regularly, which reduces the minimum size of a sample output buffer.

V. CONCLUSIONS

We have introduced the R++ model as an alternative for the LF model of the glottal pulse. It is derived from a general framework for glottal-pulse models with an exponential return phase, from which other, yet nonexisting, glottal-pulse models can be derived as well.

A psycho-acoustical comparison of stimuli generated with the R++ and the LF models showed that in some cases discrimination is possible, but that it is very unlikely that this will occur in the practical situation of speech synthesis. Even if there would be small audible differences this would not mean that one of the models would actually yield a perceptually better approximation of real speech, since both are simple models of a complex waveform and the differences between models and waveforms are much larger than the differences between the models.

We have compared the computational efficiency of the LF and the R++ model and we found that the parameter-update mechanism of R++ model is computationally more efficient and that an implementation in C of the R++ model requires less processing time. We believe, therefore, that the R++ model can serve as a useful source model in speech synthesizers.

ACKNOWLEDGMENTS

The author wishes to thank Reinier Kortekaas for his help with the setting up of the experiment.

Childers, D. G., and Lee, C. K. 1991. ''Voice quality factors: Analysis synthesis and perception,'' J. Acoust. Soc. Am. **90**, 2394–2410.

Cummings, K. E., and Clements, M. A. (**1993**). ''Application of the analysis of glottal excitation of stressed speech to speaking style modification,'' Proceedings ICASSP-93, Minneapolis, pp. 207–210.

Fant, G. (**1995**). ''The LF model revisited. Transformations and frequency domain analysis,'' Speech Transmission Laboratory Quarterly Progress Report 2-3/95, KTH, pp. 119–156.

Fant, G., Liljencrants, J., and Lin, Q. (**1985**). ''A four-parameter model of glottal flow,'' Speech Transmission Laboratory Quarterly Progress Report 4/85, KTH, pp. 1–3.

Fant, G., Kruckenberg, A., Liljencrants, J., and Båvegård, M. (**1994**). ''Voice source parameters in continuous speech. Transformation of LF parameters,'' in Proceedings of the ICSLP-94, Yokohama, pp. 1451–1454.

Gobl, C. (**1989**). ''A preliminary study of acoustic voice quality correlates,'' Speech Transmission Laboratory Quarterly Progress Report 4/89, KTH, pp. 9–22.

Green, D. A., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Holmberg, E. B., Hillman, R. E., and Perkell, J. S. 1988. ''Glottal airflow and pressure measurements for soft, normal and loud voice by male and female speakers,'' J. Acoust. Soc. Am. **84**, 511–529.

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**, 820–856.

Lin, Q. (**1990**). ''Speech Production Theory and Articulatory Speech Synthesis,'' Ph.D. thesis, Royal Institute of Technology, Stockholm.

Pierrehumbert, J. B. (**1989**). ''A preliminary study of the consequences of intonation for the voice source,'' Speech Transmission Laboratory Quarterly Progress Report 4/89, KTH.

Qi, Y., and Bi, N. 1994. ''A simplified approximation of the four-parameter LF model of voice source,'' J. Acoust. Soc. Am. **96**, 1182–1185.

Rosenberg, A., 1971. ''Effect of glottal pulse shape on the quality of natural vowels,'' J. Acoust. Soc. Am. **49**, 583–590.

Strik, H. (**1994**). ''Physiological Control and Behaviour of the Voice Source in the Production of Prosody,'' Ph.D. thesis, University of Nijmegen.