# THE EFFECT OF LOW-PASS FILTERING ON ESTIMATED VOICE SOURCE PARAMETERS

*Helmer Strik*

University of Nijmegen, Dept. of Language and Speech
P.O. Box 9103, 6500 HD Nijmegen, The Netherlands
strik@let.kun.nl, http://lands.let.kun.nl/TSpublic/strik

## ABSTRACT

Voice source parameters are often obtained by parametrizing glottal flow signals. However, before parametrization these glottal flow signals are usually low-pass filtered. As low-pass filtering changes the shape of the glottal pulses, it will also cause an error in the estimated voice source parameters. The present article presents results of our research on the effect of low-pass filtering on the estimated voice source parameters.

We will first present an evaluation method which makes it possible to study the effect of low-pass filtering in detail. The evaluation results show that low-pass filtering leads to an error in all estimated voice source parameters. However, the magnitude of the errors differs for the various voice source parameters, and also depends on the estimation method used. We will show that the errors can be reduced substantially by choosing the appropriate estimation method.

## INTRODUCTION

The technique of inverse filtering has been available for a long time now. Inverse filtering can be used to obtain an estimate of the first derivative of glottal flow ($dU_g$). Subsequently, the effect of lip radiation can be canceled by integrating $dU_g$ to obtain an estimate of true glottal flow ($U_g$). However, estimating a voice source signal (either $dU_g$ or $U_g$) is usually not enough. For many applications it is necessary to parametrize the glottal flow signals.

Before the glottal flow signals are parametrized, they are low-pass filtered at least once in all methods, viz. before A/D conversion. Often, they are low-pass filtered again after A/D conversion, usually to cancel the effects of formants that were not inverse filtered or to attenuate the noise component. However, low-pass filtering changes the shape of the glottal flow signals, and, consequently, influences the estimated voice source parameters [1, 2, 3, 4, 5, 6].

Therefore, it becomes important to determine what the effect of low-pass filtering exactly is. This was the first goal of the present study. Previously proposed methods are not optimally suited for this task (see e.g. [6]). Finding an appropriate evaluation method was the second goal of our research. The evaluation method should make it possible to determine the magnitude of the errors in the estimated voice source parameters that are due to low-pass filtering. Finally, the third goal of our study was to develop an estimation method for which these errors are small. To this end three parametrization methods are compared.

## 1. METHOD

Parametrization of $dU_g$ or $U_g$ can be done in several ways. Usually landmarks (like minima, maxima, zero crossings) are detected in the signals. Because these landmarks are estimated directly from the voice source signals, these methods will be called direct estimation methods (DE methods).

Voice source parameters can also be obtained by fitting a voice source model to the data. Because in estimation methods of this kind a model fitting procedure is used, they will be referred to as 'fit estimation' methods (FE methods). As a voice source model we use the LF model [7]. In our FE method five LF parameters ($E_e$, $t_o$, $t_p$, $t_e$, and $T_a$) are estimated for each pitch period. The FE method consists of three stages:

1. initial estimate
2. simplex search algorithm
3. Levenberg-Marquardt algorithm

As DE method we have chosen the DE method described in [4], primarily because the authors provide a fairly detailed description of their method (see especially page 765 of their article), and because with this method it is possible to estimate the LF parameters $E_e$, $t_o$, $t_p$, and $t_e$. In their method Alku & Vilkman [4] do not estimate $T_a$. Since an LF model is not complete without $T_a$, $T_a$ was estimated by fitting the second part of the LF model to the return phase of $dU_g$. Therefore, strictly speaking, only $E_e$, $t_o$, $t_p$, and $t_e$ can be said to be the result of the DE method, while $T_a$ is subsequently estimated with a fitting procedure.

In our evaluation method we first synthesize glottal flow signals. Subsequently, three parametrization methods are used to estimate the voice source parameters. Finally, the estimated voice source parameters are compared with the correct ones, i.e. those used to synthesize the glottal flow signals. As we use the LF model for the fitting procedure, it is obvious that we also used the LF model to synthesize the glottal flow signals.

The three estimation methods used in this study are pitch-synchronous. This implies that a pitch period of $dU_g$ first has to be located before it can be parametrized. Among the parameters that have to be estimated are $t_o$ and $t_c$. Because these two parameters are not known beforehand, the pitch period cannot be segmented exactly. In practice, we first locate the main excitations (i.e. $t_e$) and then use a window with a width larger than the length of the longest (expected) pitch period. Generally, the pitch period will be situated between two other pitch periods (except for UV/V and V/UV transitions). Therefore, for each experiment sequences of three equal LF pulses were used. Each time voice source parameters were estimated for the (perturbated) pulse in the middle. Another reason for not using a single glottal pulse for evaluation is that the effects of perturbations cannot always be studied by a single, isolated LF pulse.

Furthermore, LF pulses with different shapes were used. The reason is that the effect of a studied factor can depend on the shape of a pulse. Therefore, to get a general picture of the effect of that factor, the effect has to be studied for a number of pulses with different shapes. These pulses will be called the base pulses. The base pulses were obtained by using the LF model for different values of the LF parameters. The values used are based on the data given in Carlson *et al.* [8], and the data from our previous experiments [1, 2, 9, 10, 11, 12]. The shapes of the resulting 11 base pulses give a good coverage of the pulse shapes that occur during ‘normal’ running speech.

These 11 base pulses served as a starting point, and were used to generate the test pulses. To study the influence of the factor low-pass filtering, the 11 base pulses were filtered with M low-pass filters in order to generate M x 11 test pulses. Subsequently, for these test signals voice source parameters were estimated with the DE method and two FE methods. The resulting values were compared with the correct values, and the errors were calculated:

$$ERR(X) = 100\%*abs(X_{est} - X_{inp})/ X_{inp}, \text{ for } X = E_e$$

$$ERR(Y) = abs(Y_{est} - Y_{inp}), \text{ for } Y = t_o, t_p, t_e \text{ and } T_a.$$

The experiments were carried out for a number (say N) of test pulses. After calculating the errors in the estimates of the 5 LF parameters for each test pulse, the errors had to be averaged. Averaging was done by taking the median of the absolute values of the errors. The absolute values were taken because otherwise positive and negative errors could cancel each other out. In this way the average error could be small, while the individual errors are (much) larger. The median was taken because (compared to the arithmetic mean) it is less affected by outliers which are occasionally present in the estimates.

As mentioned above, the third goal of the present study was to minimize the errors due to low-pass filtering. To this end we also compared the effects of several low-pass filters. Our experiments showed that the errors caused by standard linear phase FIR filters, which are generally used as low-pass filters, are relatively large [6]. The main reason is that standard FIR filters have a ripple in their impulse response. Consequently, the low-pass filtered pulses also contain a ripple, which can have a severe influence on the parametrization. We prefer to use low-pass filters which do not have a ripple in their impulse response. We have chosen a convolution with a Blackman window, because our experiments (see e.g. [2]) revealed that this type of filter usually produces better results than other types of low-pass filters. One should thus keep in mind that for other types of low-pass filters, like e.g. the generally used linear phase FIR filters, the errors in the estimated voice source parameters will be (much) larger than the errors presented below.

The 11 base pulses were low-pass filtered by means of a convolution with a Blackman window of varying length. The length of the window was varied from 3 to 19 samples in steps of 2 samples (9 lengths). For the resulting 99 test pulses (11 base pulses x 9 window lengths) the parameters were estimated with the DE method and the FE methods. For each length of the Blackman window the results of the 11 base pulses were pooled and the median values of the absolute errors were calculated. These median values are shown in Figures 2 and 3.

## 2. RESULTS

Estimates of voice source parameters can be influenced by a large number of factors. The evaluation procedure described above makes it possible to study the effect of each individual factor and combinations of factors in detail. So far, 11 of these factors have been studied: sampling frequency, number of bits used for coding, position (shift) and amplitude ($E_e$) of the glottal pulses, $t_c$ (moment of closure), $T_0$, signal-to-noise ratio (i.e. the effect of additive noise), phase distortion (which can be caused e.g. by high-pass filtering), errors in the estimates of formant and bandwidth values during inverse filtering (which will bring about formant ripple in the estimated voice source signals), and low-pass filtering (see [5, 6, 11, 12]). Here we will focus on the effects of the factor low-pass filter.

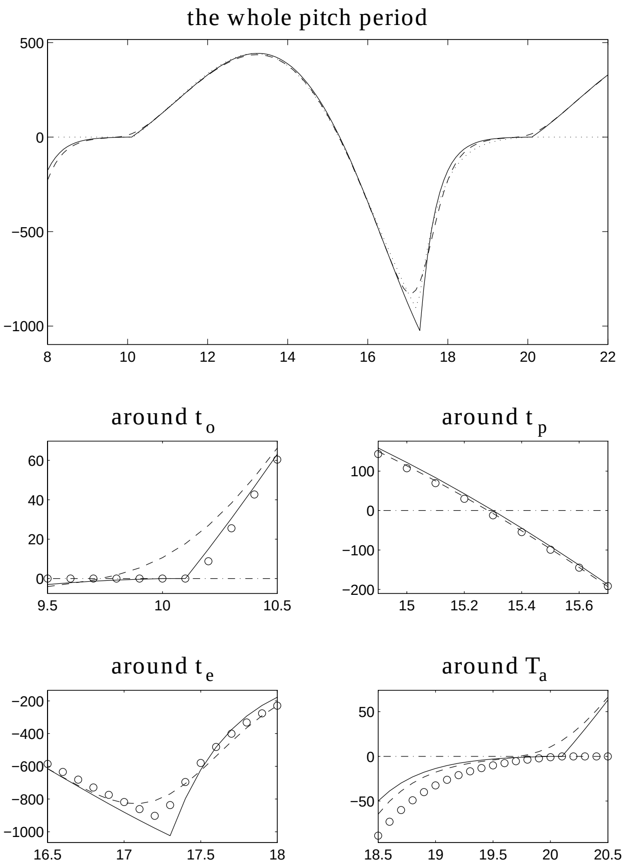

**Figure 1.** An example of a flow pulse before (solid) and after (dashed) low-pass filtering, and a fit on the low-pass filtered pulse (dotted).

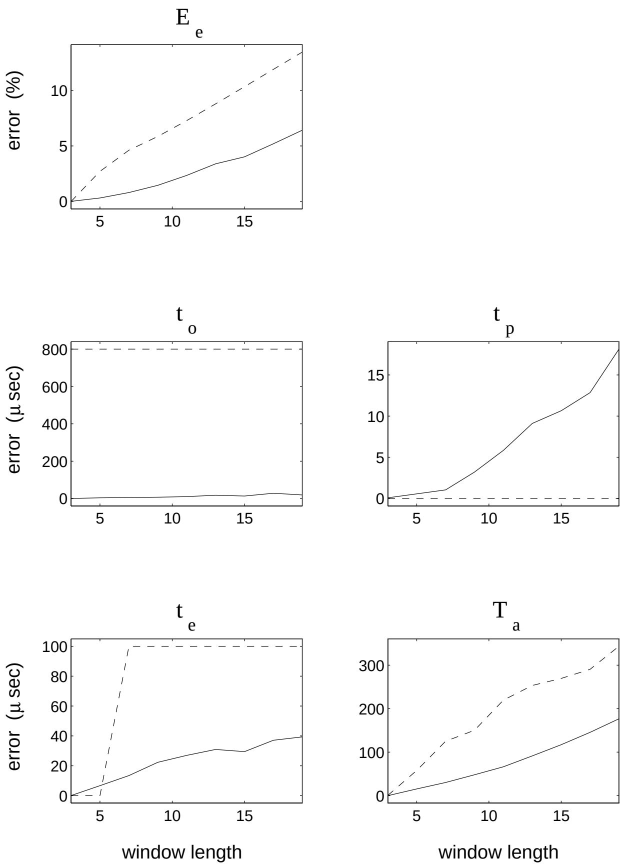

**Figure 2.** Median errors due to low-pass filtering by means of a convolution with a Blackman window whose length that varies from 3 to 19 in steps of 2. Shown are the errors for the DE method (dashed) and for the first version of the FE method (solid).

An example of the distortion of a flow pulse caused by low-pass filtering is given in Figure 1. For low-pass filtering a convolution with a 19-point Blackman window was used. Shown are a base pulse before (solid) and after (dashed) low-pass filtering, and a model fit on the low-pass filtered pulse (dotted). Besides a picture of the three signals for the whole pitch period, some details around important events are also provided.

One can see in Figure 1 that low-pass filtering does influence the shape of the pulse. From this figure one can deduce that the change in shape can have a large impact on the estimates obtained by means of a DE method. This is most clear for the estimate of $E_e$, which will generally be too small. But also the estimates of the other parameters will be affected. Low-pass filtering also affects the estimates of an FE method, but to a lesser extent.

In Figure 2 one can see that low-pass filtering affects all voice source parameters. The errors increase if the length of the Blackman window increases (i.e. if the bandwidth of the low-pass filter is reduced). Furthermore, the errors of the voice source parameters obtained with the DE method are generally larger than those obtained with the FE method. In fact, in [5] we argue that for a realistic comparison of the two methods errors due to other factors, which are always present (e.g. errors due to sampling and quantization), should be added to the errors resulting from low-pass filtering alone which are presented here. If this is done the errors for the DE method are always much larger compared to the errors of the FE method. Some details of figure 2 are further explained in [5].

In the example provided in Figure 1 the test signal is low-pass filtered. An LF model is then fitted to the low-pass filtered test pulse. This seems the most obvious way to apply low-pass filtering, and will be called the first version of the FE method. However, there is an alternative (which will be called the second version of the FE method): apart from the test pulse one could also low-pass filter the fitted LF pulse. In that case, test pulse and fitted LF pulse are altered in a similar fashion. In this way we hope to achieve that the error in the estimated parameters (which is due to low-pass filtering) will be smaller than when only the test pulses are low-pass filtered. It is obvious that the same trick cannot be used in a DE method, because in this case the parameters are calculated directly from the (low-pass filtered) signal.

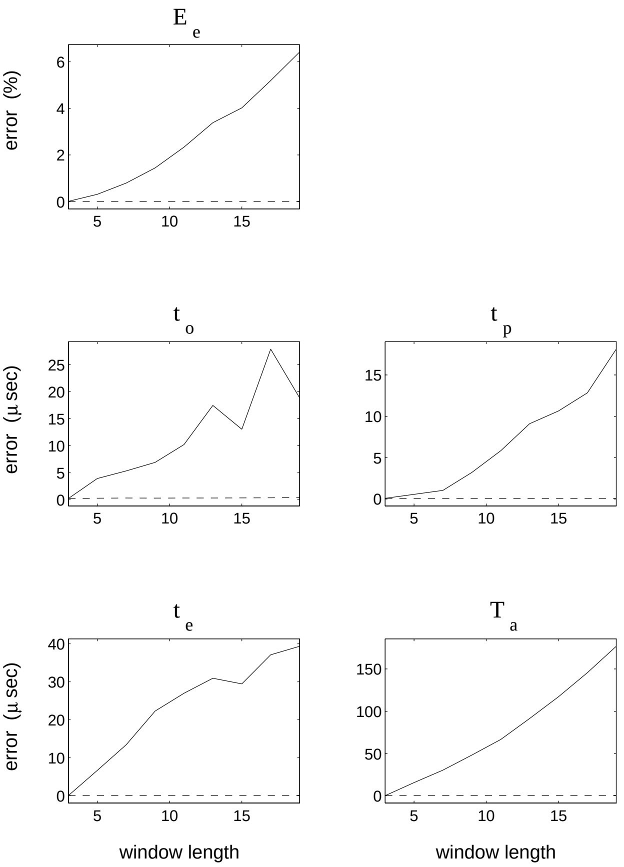

**Figure 3.** Median errors due to low-pass filtering by means of a convolution with a Blackman window whose length varies from 3 to 19 in steps of 2. Shown are the errors for the first (solid) and the second (dashed) version of the FE method.

In Figure 3 the results of the two versions of the FE method are compared, i.e. the first version, in which only the test pulses are low-pass filtered (solid lines), and the second version, in which both test pulses and fitted LF pulses are low-pass filtered (dashed lines). Clearly, the errors for the second version are much smaller. The errors are not zero, as may seem to be the case from Figure 3, but they are extremely small. The largest error observed in the time parameters is 1 msec., and the errors in $E_e$ are always smaller than 0.03%.

## CONCLUSIONS

With the evaluation method proposed above it becomes possible to make an accurate study of various factors that affect the estimated voice source parameters. The results for the factor low-pass filtering are presented in the current article. These results show that low-pass filtering causes an error in all estimated voice source parameters. The magnitude of the errors differs for the various voice source parameters, and between estimation methods. The largest errors were found for DE methods, which are used most often in practice. By fitting a voice source model to the data the errors can be reduced. A further drastic reduction in the error can be obtained if the fitted voice source model is filtered with the same low-pass filter as used to filter the glottal flow signals.

## REFERENCES

[1] Strik, H., J. Jansen and L. Boves (1992), ‘Comparing methods for automatic extraction of voice source parameters from continuous speech’, *Proc. ICSLP ’92*, Banff, 121-124.

[2] Strik, H., B. Cranen and L. Boves (1993), ‘Fitting a LF-model to inverse filter signals’, *Proc. EUROSPEECH ‘93*, Berlin, 103-106.

[3] Perkell, J. S., Hillman, R. E., and Holmberg, E. B. (1994), ‘Group differences in measures of voice production and revised values of maximum airflow declination rate’, *J. Acoust. Soc. Am*. 96, 695-698.

[4] Alku, P. and E. Vilkman (1995), ‘Effects of bandwidth on glottal airflow waveforms estimated by inverse filtering’, *J. Acoust. Soc. Am.*, 98, 763-767.

[5] Strik, H. (1996), ‘Testing two automatic methods for estimation of voice source parameters’, *Proc. of the Department of Language & Speech*, Vol. 19, pp. 105-127, Nijmegen, The Netherlands, 1996.

[6] Strik, H. (1996), ‘Comments on "Effects of bandwidth on glottal airflow waveforms estimated by inverse filtering" [J. Acoust. Soc. Am. 98, 763-767 (1995)]’, *J. Acoust. Soc. Am*., 100, 1246-1249.

[7] Fant, G., J. Liljencrants and Q. Lin (1985), ‘A four parameter model of glottal flow’, *Speech Transmission Laboratory, Q. Prog. Status Rep.*, Royal Institute of Technology, Stockholm, 4/1985, 1-13.

[8] Carlson, R., G. Fant, C. Gobl, B. Granstrom, I. Karlsson and Q. Lin (1989), ‘Voice source rules for text-to-speech synthesis’, *Proc. ICASSP’89*, 223-226.

[9] Strik, H. and L. Boves (1992), ‘Control of fundamental frequency, intensity and voice quality in speech’, *Journal of Phonetics*, 20, 15-25.

[10] Strik, H. and L. Boves (1992), ‘On the relation between voice source parameters and prosodic features in connected speech’, *Speech Communication*, 11, 167-174.

[11] Strik, H. and L. Boves (1994), ‘Automatic estimation of voice source parameters’, *Proc. ICSLP ‘94*, Yokohama, Japan, pp. 155-158.

[12] Strik, H. (1994), *Physiological control and behaviour of the voice source in the production of prosody*, PhD dissertation, University of Nijmegen.

At http://lands.let.kun.nl/TSpublic/strik PostScript and ASCII versions of most of my publications can be found.