ANALYSIS OF GLOTTAL WAVEFORM IN DIFFERENT PHONATION TYPES USING THE NEW IAIF-METHOD

P. Alku[1], E. Vilkman[2], U.K. Laine[1]

1: Helsinki University of Technology, FINLAND
2: University of Oulu, FINLAND

ABSTRACT

A new glottal wave analysis method, IAIF (Iterative Adaptive Inverse Filtering), is presented. In this algorithm the effect of glottal pulses to the speech spectrum is first estimated with an iterative procedure. The model for the vocal tract is then computed using linear prediction. Finally, the effects of the vocal tract and lip radiation are cancelled by inverse filtering. The IAIF-method was tested using synthetic and natural vowels of three different phonation types. The new algorithm was able to yield a fairly accurate estimate for the glottal excitation excluding the case of very pressed phonation that was partly distorted by a formant ripple.

1. INTRODUCTION

Many different methods have been developed during the last fourty years in order to estimate the source of voiced speech, the glottal pulseform. One of the most popular techniques that is applied in the analysis of the glottal excitation is inverse filtering. Although good results have been obtained with this method it has some drawbacks. For instance, the transfer function of the inverse filter is often adjusted manually. Hence, the final result is very much dependent on the subjective criteria applied by the researcher [e.g. 3]. Another drawback that is characteristic especially to the closed phase covariance method is that the analysis works properly only for phonation types with a sufficiently long glottal closed phase.

In this paper a new glottal wave analysis method, IAIF (Iterative Adaptive Inverse Filtering), is presented. The method represents further development of the AIF-method, which has been presented earlier [1]. Therefore a brief description of the algorithm will be given in section 2. The performance of the IAIF-method in the estimation of the glottal excitation is discussed in section 3 using both synthetic and natural utterances.

2. METHOD

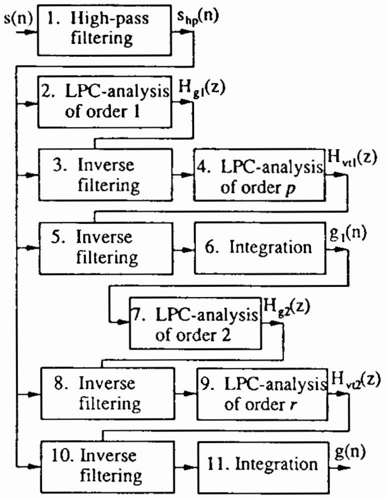
The IAIF-method is based on a speech production model that consists of three separated processes: the glottal excitation, the vocal tract and the lip radiation effect. The model is assumed to be linear and the interaction between the three parts is considered to be negligible. The vocal tract is modeled with an all-pole filter. The last process of the model, the lip radiation effect, is modeled with a differentiator.

The block diagram of the IAIF-method is shown in Fig. 1. The speech signal to be analysed is denoted s(n) and the result, the estimate for the glottal excitation, is denoted g(n). The first iteration consists of the blocks numbered from 2 to 6 and the second iteration of the blocks numbered from 7 to 11. The purpose of each of the blocks is described as follows.

<u>Block no. 1:</u>
In order to remove undesirable fluctuations of the output of the integrator signal s(n) has to be high-pass filtered. The high-pass filter is a linear phase FIR having 511 coefficients and a cut-off frequency of 20 Hz.

<u>Block no. 2:</u>
The effect of the glottis to the speech spectrum is preliminarily estimated by first order LPC-analysis.
<u>Block no. 3:</u>
The estimated glottal contribution is eliminated by filtering $s_{hp}(n)$ through $H_{g1}(z)$.
<u>Block no. 4:</u>
The first estimate for the vocal tract is computed by applying LPC-analysis to the output of the previous block.
<u>Block no. 5:</u>
The effect of the vocal tract is eliminated from signal $s_{hp}(n)$ by inverse filtering.
<u>Block no. 6:</u>
The first estimate for the glottal excitation $g_1(n)$, is obtained by cancelling the lip radiation effect by integrating.
<u>Block no. 7:</u>
The second iteration starts by computing a new estimate for the effect of the glottis to the speech spectrum. This time second order LPC-analysis is used. The signal from which the glottal contribution is estimated is $g_1(n)$.
<u>Block no. 8:</u>
The effect of the estimated glottal contribution is eliminated.
<u>Block no. 9:</u>
The final model for the vocal tract is obtained by applying LPC-analysis of order *r* to the output of the previous block.
<u>Block no. 10:</u>
The effect of the vocal tract is eliminated from speech by filtering $s_{hp}(n)$ through $H_{vt2}(z)$.
<u>Block no. 11:</u>
The result, g(n), is obtained by cancelling the lip radiation effect by integrating the output of block no. 10.

The results discussed in this paper are based on the implementation of the IAIF-algorithm on a Symbolics Lisp-machine.



Transfer functions of the filters are:

$$H_{g1}(z) = 1 + az^{-1} \qquad H_{vt1}(z) = 1+\sum_{k=1}^{p} a(k)z^{-k}$$

$$H_{g2}(z) = 1 + bz^{-1} + cz^{-2} \qquad H_{vt2}(z) = 1+\sum_{k=1}^{r} b(k)z^{-k}$$

Fig. 1. Block diagram of the IAIF-method

3. RESULTS

3.1. Synthetic vowels

In order to verify the performance of the IAIF-method the new algorithm was first tested with synthetic speech. Synthetic vowels were created using a procedure described in [4]. The vocal tract was modeled with an eighth order all-pole filter and the lip radiation effect with a differentiator. The shape of the vocal tract transfer function corresponded to the vowel /a/. The signal bandwidth was 4 kHz. As the synthetic source signal we used a glottal pulse model described in [2]. Three different phonation types, breathy, normal and pressed, were simulated by changing the shape of the synthetic excitation waveform. Two different values for the pitch period, corresponding to male and female speakers, were used in the synthesis procedure.

The IAIF-analysis was computed for all the signals using a block length of 256 samples (32 ms). The orders of LPC-analysis corresponding to modeling of the vocal tract (parameters *p* and *r* of blocks no. 4 and 9 in Fig. 1) were chosen to be equal. This value was varied from 8 to 12 by a step of two.

When synthetic male phonation was analysed the IAIF-method yielded a result that was very close to the original source signal. In the case of breathy and normal phonation similarity between the original source and the waveform given by the IAIF-method was almost exact without dependence on parameter p. A typical result is shown in Fig. 2. In the case of pressed phonation the waveform obtained by the IAIF-method was partly distorted by a ripple component when the value of p was equal to 8 i.e. to the order of the all-pole vocal tract. However, by increasing the order of p to be equal to 12 the ripple component disappeared.

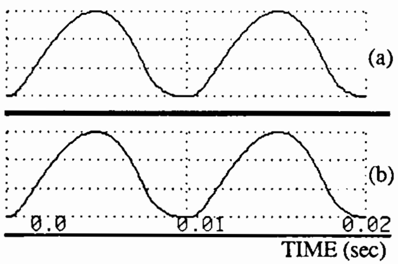


Fig. 2. Analysis of a synthetic male vowel of breathy phonation
(a): Original synthetic glottal source
(b): Glottal wave estimate given by the IAIF-method (p=8)

When synthetic female utterances were analysed the results were not so good as for male voices. In the case of breathy phonation the IAIF-method gave a waveform that was similar to the synthetic source signal. However, for normal and in particular for pressed phonation types the result given by the new algorithm was partly distorted by a ripple component. This results from the spectrum of the glottal excitation which in the case of pressed phonation comprises more high frequency components than in the case of breathy or normal phonation. In the case of female voice the source spectrum is also characterized by a sparse harmonic structure. Hence, LPC-analysis (block no. 9 of Fig. 1) gives a vocal tract filter, where the formants, especially F1, are moved from their original positions because of the harmonics of the source spectrum. Thus, a small formant ripple will be present in the glottal wave estimate after inverse filtering and integration.

3.2. Natural vowels

The IAIF-method was used in the glottal wave analysis of sustained phonation by studying utterances that were produced by one female and one male speaker. Both of the subjects were of healthy voice. The speakers were asked to produce the vowel /a/ using breathy, normal and pressed phonation. The recording was done in an anechoic chamber using a condenser microphone (Brüel&Kjær 4134). The speech material was A/D-converted with Sony PCM-F1 and stored on a video cassette using Sony SL-F1E. The bandwidth of the signals was downsampled to 4 kHz.

In the case of male voice the results were of reliable shapes for breathy and normal phonation types. The glottal waveform corresponding to pressed phonation was partly distorted by a formant ripple. Fig. 3 shows the obtained glottal pulseforms for all the three phonation types.

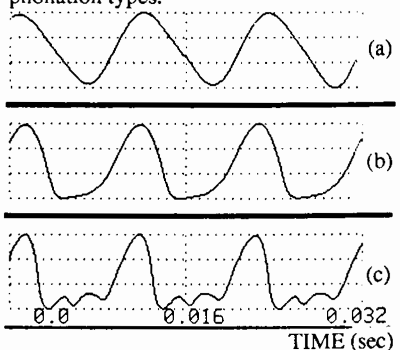


Fig. 3. Glottal wave estimates given by the IAIF-method (natural male voice, p=12)
(a): Breathy phonation
(b): Normal phonation
(c): Pressed phonation

The analysis of female voice yielded results that were, quite surprisingly, free from formant ripple for all the three phonation types. The waveform of breathy phonation was of a very smooth shape. No clear closed phase could be distinguished. The time instant of the maximum glottal opening occurred approximately in the middle of the glottal cycle. In the case of normal phonation the time instant of the maximum opening was moved to the point that corresponds to 70 % of the length of the pitch period. The waveform of pressed phonation was the only one with a clear closed phase.

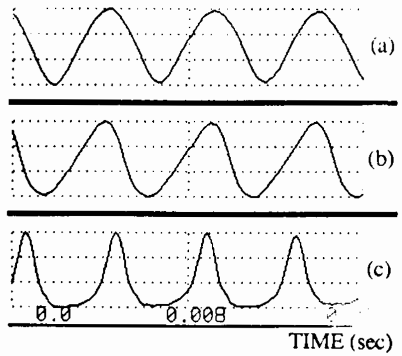


Fig. 4. Glottal wave estimates given by the IAIF-method (natural female voice, p=12)
(a): Breathy phonation
(b): Normal phonation
(c): Pressed phonation

4. DISCUSSION

In this paper a new glottal wave analysis tool, the IAIF-method, was presented. The identification of the different processes of the human speech production mechanism is done in the new algorithm using a frequency domain approach. The average glottal contribution to the speech spectrum is first estimated with an iterative procedure. The vocal tract is then identified by LPC-analysis. The estimate for the glottal excitation is finally obtained by cancelling the effects of the vocal tract and lip radiation by inverse filtering.

The new IAIF-algorithm was applied in this study for the glottal wave analysis using three different phonation types. The results obtained are well in line with those reported using other methods [e.g. 3].

In the case of male voice both synthetic and natural utterances gave the same result: breathy and normal phonation can be analysed accurately whereas pressed phonation is partly distorted by a formant ripple. The reason for distortion with the IAIF-method was obviously the poor estimation of the first formant, which comes from the contribution of the source spectrum. For synthetic female voices, especially in the case of pressed phonation, distortion was largest. However, in general, excluding the very pressed phonation type, the source spectrum of natural female phonation decays so fast that the first formant can be modeled properly. This explains why the analysis results obtained from the utterances of the female subject were of reliable shapes with no formant ripple.

The IAIF-method has proved to be a promising tool for glottal wave analysis. The main advantage of the new algorithm is that it is automatic. Hence, the glottal pulseform can be obtained without manual interference by the investigator. Further studies are needed to compare the IAIF-technique with traditional methods as well as to reveal whether it can be used for analysis of connected speech. Also the real-time implementation of the algorithm using the TMS320C30-signal processor is under development.

References:

[1] Alku, P., Vilkman, E., Laine, U.K. (1990), "A comparison of EGG and a new inverse filtering method in phonation change from breathy to normal," Proc. Int. Conf. on Spoken Language Processing, pp. 197-200.

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

[3] Gauffin, J., Sundberg, J. (1989) "Spectral correlates of glottal voice source waveform characteristics," J. Speech and Hearing Research, Vol. 32, pp. 556-565.

[4] Gold, B., Rabiner, L.R. (1968), "Analysis of digital and analog formant synthesizers," IEEE Trans. Audio and Electroacoustics, Vol. 16, pp. 81-94.

[5] Wong, D.Y., Markel, J.D., Gray, A.H., Jr. (1979), "Least squares glottal inverse filtering from acoustic speech waveforms," IEEE Trans. Acoust., Speech, Signal Processing, Vol. 27, pp. 350-355.