2nd European Conference on
Speech Communication and Technology
EUROSPEECH '91
Genova, Italy, September 24-26, 1991

GLOTTAL WAVE ANALYSIS WITH PITCH SYNCHRONOUS ITERATIVE ADAPTIVE INVERSE FILTERING

Paavo Alku

Helsinki University of Technology, Acoustics Laboratory
Otakaari 5 A, SF-02150 Espoo, FINLAND

ABSTRACT

A new glottal wave analysis method, Pitch Synchronous Iterative Adaptive Inverse Filtering (PSIAIF), is presented. The algorithm takes advantage of a previously developed glottal wave analysis technique, Iterative Adaptive Inverse Filtering (IAIF). The basic idea in the IAIF-algorithm is that the average effect of the glottal excitation to the speech spectrum is first estimated using an adaptive low-order all-pole filter. The estimate for the vocal tract is then computed by applying LPC-analysis to the signal from which the estimated glottal contribution was eliminated. The glottal excitation is obtained by eliminating the effects of the vocal tract and lip radiation by inverse filtering. The new PSIAIF-method applies first the IAIF-analysis described above pitch asynchronously to a speech signal. The resulting glottal wave estimate is used in order to determine positions of frames for pitch synchronous IAIF-analysis. The final estimate for the glottal pulseform is obtained by analysing the original speech signal by the IAIF-algorithm using frames that span speech samples between consecutive maximal glottal openings. Preliminary results show that the PSIAIF-technique is able to give a fairly reliable estimate for the glottal excitation.

1. INTRODUCTION

The excitation of the vocal tract, the glottal waveform, has an important role in speech production. Many different methods have been developed during the past years in order to obtain reliable information about the behavior of this source signal. Among the glottal wave analysis techniques one of the most popular is inverse filtering. In general this method is believed to estimate the glottal excitation properly. However, some inverse filtering techniques are not able to analyse speech where the length of the closed period of the glottal cycle is short. Also the manual adjustment of inverse filtering parameters may have an effect that the results are dependent on the subjective criteria applied by the researcher.

In this article a new glottal wave analysis method, Pitch Synchronous Iterative Adaptive Inverse Filtering (PSIAIF), is presented. The new algorithm is, in comparison to many older inverse filtering techniques, fully automatic and non-invasive. According to preliminary studies the new method is able to give fairly good estimates for the glottal pulseform.

The PSIAIF-algorithm is a sequel to two previously developed procedures, Adaptive Inverse Filtering (AIF) [1] and Iterative Adaptive Inverse Filtering (IAIF) [2]. The main purpose of this paper, the description of the principle of the new algorithm, is discussed in section 2. The preliminary results that were obtained when the new algorithm was used in the analysis of natural speech are discussed in section 3.

2. METHOD

2.1 General

The PSIAIF-algorithm is based on the application of the IAIF-algorithm, which has found to be a promising method in the estimation of the glottal excitation [2]. In the new PSIAIF-method the computation of the glottal waveform is performed in two steps. In the first step the estimate for the glottal pulseform is computed pitch asynchronously with the IAIF-method. The resulting glottal pulseform is then analysed in order to determine the positions of frames that span individual pitch periods. In the second step of the PSIAIF-algorithm the original speech signal is analysed pitch synchronously with the IAIF-algorithm using frames that were determined by the first step.

The role of the IAIF-algorithm is essential in the PSIAIF-method. Therefore the structure of the IAIF-algorithm is first discussed.

2.2 Structure of the IAIF-algorithm

The IAIF-method is based on a separated linear speech production model. The interaction between the three different parts of the model, the glottal excitation, the vocal tract and the lip radiation effect, is considered to be negligible. The transfer function of the vocal tract is estimated by the IAIF-algorithm using an adaptive all-pole filter. The lip radiation effect is modeled with a fixed differentiator.

Computation of the glottal excitation is based on the idea that the average effect of the glottal source to the speech spectrum is first estimated with an adaptive low order all-pole filter. The model for the vocal tract is obtained by applying linear predictive coding (LPC) to the signal from which the estimated glottal contribution was eliminated.

The IAIF-algorithm consists of two iterations. In the first round the glottal contribution is estimated by applying first order linear prediction to the original speech signal. The result of this iteration, the first estimate for the glottal excitation, is used in the beginning of the second iteration in order to obtain a better estimate for the effect of the glottal source to the speech spectrum.

The block diagram of the IAIF-algorithm is shown in Fig. 1. The first iteration consists of blocks numbered from 1 to 5 and the second iteration of the blocks numbered from 6 to 10. The purpose of the blocks is decribed as follows:

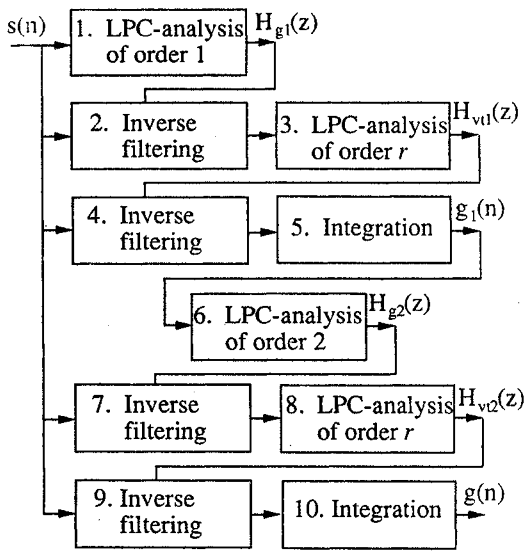
Block no. 1:
The effect of the glottal source to the speech spectrum is preliminarily estimated by first order LPC-analysis.
Block no. 2:
The estimated glottal contribution is eliminated by filtering s(n) through $H_{g1}(z)$.
Block no. 3:
The first estimate for the vocal tract is computed by applying LPC-analysis to the output of the previous block.
Block no. 4:
The effect of the vocal tract is eliminated from signal s(n) by inverse filtering.
Block no. 5:
The first estimate for the glottal excitation, $g_1(n)$, is obtained by cancelling the lip radiation effect by integrating.
Block no. 6:
The second iteration begins by computing a new estimate for the effect of the glottal source to the speech spectrum. This time second order LPC-analysis is used. The signal from which the glottal contribution is estimated is $g_1(n)$.
Block no. 7:
The effect of the estimated glottal contribution is eliminated.
Block no. 8:
The final model for the vocal tract is obtained by applying LPC-analysis of order *r* to the output of the previous block.
Block no. 9:
The effect of the vocal tract is eliminated from speech by filtering s(n) through $H_{vt2}(z)$.
Block no. 10:
The result, g(n), is obtained by cancelling the lip radiation effect by integrating the output of block no. 9.

2.3 Structure of the PSIAIF-algorithm

The block diagram of the PSIAIF-algorithm is shown in Fig. 2. The speech signal to be analysed is denoted s(n). The result, the estimate for the glottal excitation, is denoted g(n).

As a first stage of the algorithm the speech signal is high-pass filtered in order to remove undesirable fluctuations of the resulting glottal pulseform. As a high-pass filter we have used a linear phase FIR, that has 2047 coefficients and a cut-off frequency of 30 Hz.

The high-pass filtered speech signal, $s_{hp}(n)$, is analysed first (block no. 2 of Fig. 2) with the IAIF-algorithm as described in



Transfer functions of the filters are:

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

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

Fig. 1 Block diagram of the IAIF-algorithm

section 2.2. The length of the analysis frame is fixed and long enough to span several pitch periods. As a result of this pitch asynchronous stage a glottal wave estimate, denoted $g_{pa}(n)$, is obtained.

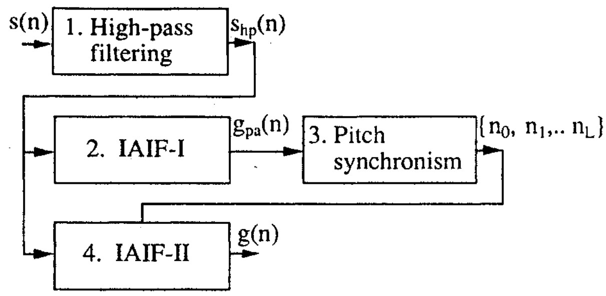


Fig. 2 Block diagram of the PSIAIF-method

In the next stage of the algorithm (block no. 3) positions of frames for the pitch synchronous IAIF-analysis are computed with the help of signal $g_{pa}(n)$. A pitch synchronous analysis frame, as shown in Fig. 3, is determined as a time interval between two consecutive maximal glottal openings. This choice for the position of the frame is justified as follows. First, by using both synthetic and natural speech it was found that the best results were obtained when the position of pitch synchronous frame was adjusted to begin from the time instant of maximal glottal opening. Second, determination of the time instants of maximal glottal openings can be computed reliably and easily from the signal $g_{pa}(n)$ without any manual interference.

Time instants of maximal glottal openings (n_0, n_1, n_2 in Fig. 3) are determined as follows. The length of the pitch period, denoted M, is first estimated as the index of the maximum of the autocorrelation function of $g_{pa}(n)$ between time lags of 2.5 ms and 15ms. The time index of maximal glottal opening, n_1 of Fig. 3, for example, is then determined by searching for the maximal amplitude of $g_{pa}(n)$ during the time that spans from $n_0+0.5\cdot M$ to $n_0+1.5\cdot M$, where n_0 is a previously determined index of maximal glottal opening. The procedure is repeated until all the indices of maximal glottal openings, denoted $\{n_0, n_1, .. n_L\}$ in Fig. 2, that occur within one pitch asynchronous frame of $g_{pa}(n)$ are obtained.

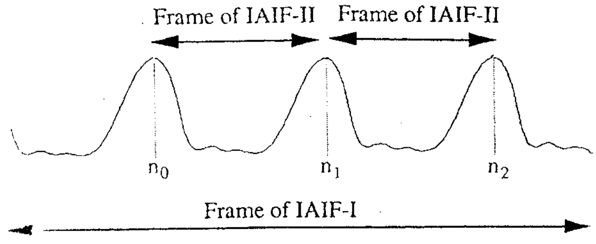


Fig. 3 Determination of frames for the pitch synchronous analysis from signal $g_{pa}(n)$

Finally (stage no. 4 of Fig. 2) the high-pass filtered speech signal is once again analysed with the IAIF-algorithm. However, this time the analysis is computed in frames whose borders are determined by the indices of consecutive maximal glottal openings i.e. $\{n_0, n_1, .. n_L\}$. The output of stage no. 4 forms the final estimate for the glottal excitation.

3. EXPERIMENTS

3.1 Speech material

In this preliminary study the PSIAIF-algorithm was applied in the glottal wave analysis of natural speech that was produced by two female and two male speakers. All the subjects were of healthy voices. The speakers were asked to produce four words of the Finnish language using normal phonation. Each of the words ("ei" /ei/, "ou" /ou/, "ai" /ai/, and "ui" /ui/) consisted of two vowels.

The speech material was recorded in an anechoic chamber using a condenser microphone (Brüel&Kjær 4133). The signals were A/D-converted using Sony PCM-F1 and stored on a video cassette using Sony SL-F1E. The data was transferred to a Symbolics Lisp-machine where the PSIAIF-algorithm has been implemented. The bandwidth of the signals was decreased to 4 kHz.

All the signals were analysed using a block length of 256 samples (32 ms) in the pitch asynchronous IAIF-block (stage no. 2 of Fig. 2). Autocorrelation criterion was used together with Hamming-windowing in all the LPC-analysis that were needed to be computed in the PSIAIF-algorithm. The order of LPC-analysis concerning modeling of the vocal tract (parameter *r* in blocks no. 3 and 8 of Fig. 1) was 12. The energy of each of the computed glottal pulses was scaled to unity.

From the author's point of view one of the most important reasons for this study was to find out if the pitch synchronous analysis applied by the PSIAIF-algorithm gave any improvements compared to the previously developed IAIF-method. The performances of the two methods can be compared easily because the result of the IAIF-algorithm (i.e. signal $g_{pa}(n)$ of Fig. 2) can be obtained as a byproduct of the PSIAIF-structure.

3.2 Results

When male speech was analysed with the PSIAIF-method the obtained pulseforms were of reliable shapes. Fig. 4 and 5 show typical results. The closed phase of the glottal pulseform was for some of the male signals partly distorted by a formant ripple. However, this distortion was clearly smaller in the case of PSIAIF-analysis when compared to pitch asynchronous IAIF-analysis. Especially when analysing that part of the signal where a vowel was gliding from one to another the PSIAIF-method gave a more reliable result. This results from the fact that the transfer function of the vocal tract is estimated more accurately by consecutive pitch synchronously computed IAIF-analysis (block no. 4 of Fig. 2) than by one pitch asynchronously computed IAIF-algorithm (block no. 2 of Fig. 2).

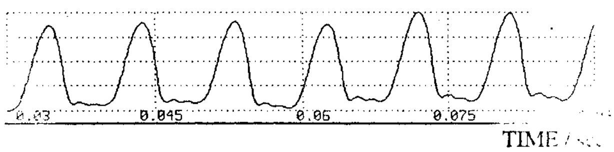


Fig. 4 Glottal wave estimate for male speech (vowel /a/ in word "ai")

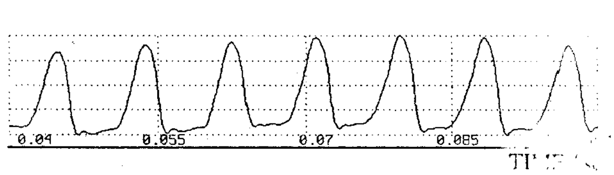


Fig. 5 Glottal wave estimate for male speech (vowel /e/ in word "ei")

When female speech was analysed the obtained pulseforms corresponded also quite well with the results that have been obtained by other methods. The glottal wave estimates were of smooth shapes and the length of the closed phase was quite short. Typical results are shown in Fig. 6 and 7. If the length of the pitch period is very small the number of samples that is needed for LPC-analysis of the pitch synchronous IAIF-analysis is not sufficient. Hence, in the analysis of signals of high F0 it is recommended to rather apply the pitch asynchronous IAIF-algorithm.

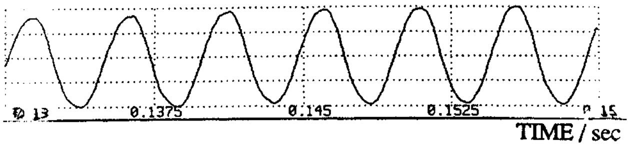


Fig. 6 Glottal wave estimate for female speech (vowel /u/ in word "ui")

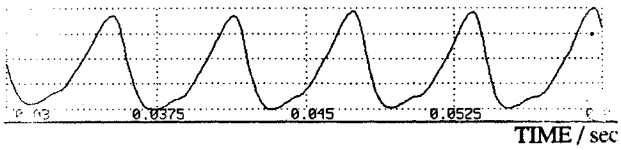


Fig. 7 Glottal wave estimate for female speech (vowel /o/ in word "ou")

4. SUMMARY AND CONCLUSIONS

In this paper a new glottal wave analysis method, the PSIAIF-algorithm, was presented. The new method takes advantage of the previously developed glottal wave analysis method, the IAIF-algorithm. The IAIF-method is based on the idea that the vocal tract is estimated by first eliminating from the speech spectrum the average effect of the glottal excitation. The effect of the glottal source to the speech spectrum is estimated by a low-order all-pole filter that is computed with an iterative algorithm. The PSIAIF-algorithm computes the glottal wave estimate one pulse at a time by applying the IAIF-technique. The position of pitch synchronous frames are determined from pitch asynchronously computed glottal wave analysis.

The preliminary results that were obtained when the PSIAIF-method was used in the glottal wave analysis of natural speech are promising. The obtained waveforms were of reliable shapes for male as well as for female speech signals that were studied. The pitch synchronous analysis applied by the PSIAIF-method improved the results compared to those that were obtained by the pitch asynchronous IAIF-method. Since the model of the vocal tract is updated once every pitch period by the PSIAIF-algorithm the new method is able to estimate the glottal excitation more reliably in the case when a vowel rapidly changes from one to another. The pitch synchronous computation improved the results especially in the analysis of vowels of low first formant. The reason is that the estimation of the vocal tract transfer function with LPC-analysis can be affected by the harmonic structure of the source spectrum if the analysis is computed during several pitch periods. When computation is done pitch synchronously the harmonic structure is absent. However, it should be noted that if the length of the pitch period is small the performance of the PSIAIF-algorithm may decrease. This happens because the estimation of the vocal tract with LPC-analysis deteriorates due to an unsufficient number of speech samples.

REFERENCES:

[1] Alku, P., Vilkman, E., Laine, U.K., "A comparison of EGG and a new inverse filtering method in phonation change from breathy to normal", Proc. Int. Conf. on Spoken Language Processing-90, pp. 197-200.

[2] Alku, P., Vilkman, E., Laine, U.K., "Analysis of glottal waveform in different phonation types using the new IAIF-method," to be published in Proc. of the XIIth Int. Congress of Phonetic Sciences-91.

[3] Hunt, M.J., "Studies of glottal excitation using inverse filtering and an electroglottograph", Proc. of the XIth Int. Congress of Phonetic Sciences-87, Vol. 3, pp. 23-26.

[4] Wong, D.Y., Markel, J.D., Gray, A.H., Jr., "Least squares glottal inverse filtering from the acoustic speech waveform", IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-27, No. 4, pp. 350-355, Aug. 1979.