Speech Communication 11 (1992) 109–118
North-Holland

Glottal wave analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering

Paavo Alku
Helsinki University of Technology, Acoustics Laboratory, Otakaari 5A, SF-02150 Espoo, Finland

Received 26 September 1991
Revised 23 January 1992

Abstract. A new glottal wave analysis method, Pitch Synchronous Iterative Adaptive Inverse Filtering (PSIAIF) is presented. The algorithm is based on a previously developed method, Iterative Adaptive Inverse Filtering (IAIF). In the IAIF-method the glottal contribution to the speech spectrum is first estimated with an iterative structure. The vocal tract transfer function is modeled after eliminating the average glottal contribution. The glottal excitation is obtained by cancelling the effects of the vocal tract and lip radiation by inverse filtering. In the new PSIAIF-method the glottal pulseform is computed by applying the IAIF-algorithm twice to the same signal. The first IAIF-analysis gives as a result a glottal excitation that spans over several pitch periods. This pulseform is used in order to determine positions and lengths of frames for the pitch synchronous analysis. The final result is obtained by analysing the original speech signal with the IAIF-algorithm one fundamental period at a time. The PSIAIF-algorithm was applied in glottal wave analysis using both synthetic and natural vowels. The results show that the method is able to give a fairly accurate estimate for the glottal flow excluding the analysis of vowels with a low first formant that are produced with a pressed phonation type.

Zusammenfassung. Im vorliegenden Artikel wird ein neues Verfahren zur Analyse des Glottissignals vorgestellt, das Pitch Synchronous Iterative Adaptive Inverse Filtering (PSIAIF). Der Algorithmus basiert auf einer vorher entwickelten Methode (IAIF). Bei der IAIF-Methode wird zunächst der glottale Beitrag zum Sprachspektrum mit einem iterativen Verfahren geschätzt. Die Übertragungsfunktion des Vokaltraktes wird anschließend nach Eliminierung des mittleren Glottalbeitrags modelliert. Die Anregung der Glottis wird ermittelt, indem die Einflüsse von Vokaltrakt und Lippenabstrahlung mittels inverser Filterung beseitigt werden. In der neuen PSIAIF-Methode wird die glottale Pulsform dadurch bestimmt, daß der IAIF-Algorithmus zweimal auf dasselbe Signal angewendet wird. Die erste IAIF-Analyse liefert als Ergebnis die Glottisanregung über mehrere Grundperioden. Diese Pulsform wird dann dazu benutzt, Rahmenpositionen und -längen für die grundperiodensynchrone Analyse zu ermitteln. Das endgültige Ergebnis erhält man, indem das ursprüngliche Sprachsignal in jeweils einer Periode mit dem IAIF-Algorithmus analysiert wird. Der PSIAIF-Algorithmus wurde sowohl unter Verwendung von synthetischen als auch natürlichen Vokalen erprobt. Die Ergebnisse zeigen, daß das Verfahren dazu in der Lage ist, eine hinreichend genaue Schätzung des Glottissignals anzugeben, mit Ausnahme von Vokalen mit einer niedrigen Frequenz des ersten Formanten und welche durch eine gepresste Aussprache erzeugt werden.

Résumé. On présente une nouvelle méthode d'analyse du flux glottique: le PSIAIF (Pitch Synchronous Iterative Adaptive Inverse Filtering). Cet algorithme se base sur une méthode (IAIF) développée précédemment. La contribution glottique totale au spectre de la parole y était tout d'abord évaluée itérativement. La fonction de transfert du conduit vocal était obtenue après élimination de la contribution glottique moyenne tandis que l'excitation glottique l'était en annulant les effets du conduit vocal et de la radiation labiale par filtrage inverse. Dans la nouvelle méthode, l'onde glottique est calculée en appliquant deux fois l'algorithme IAIF au même signal. La première analyse donne une estimation de l'excitation glottique qui s'étend sur plusieurs périodes. L'onde ainsi obtenue est utilisée ensuite pour déterminer les positions et les longueurs des fenêtres d'analyse synchronisées. Pour obtenir le résultat final, il ne reste plus qu'à analyser le signal original de la parole, période fondamentale par période fondamentale, avec l'algorithme IAIF. L'algorithme PSIAIF a été appliqué à l'analyse du signal glottique, dans le cas de voyelles naturelles et synthétiques. Les résultats montrent que la méthode est capable de fournir une estimation relativement précise de flux glottique, si l'on exclut l'analyse des voyelles à premier formant bas produites par un type de phonation pressée.

Keywords. Glottal wave analysis; inverse filtering.

1. Introduction

The air flow through the vibrating vocal folds, the glottal pulseform, serves as the excitation of the speech production mechanism. Hence, in the analysis of speech the role of the glottal flow is of great importance. To compute a reliable estimate for the glottal waveform has been a target of speech research for several decades. Over the years many different methods have been developed in order to obtain information from the source signal. Unfortunately, an analysis method that is both easy to use and which gives accurate results for a large variety of speech signals has not yet been developed.

A widely used technique in the estimation of the glottal excitation is inverse filtering. Even though many different versions of this approach have been developed they are all based on the same idea according to which the glottal pulseform is obtained by cancelling the effect of formants from speech. The model for the vocal tract needs to be computed first and the effect of formants is then cancelled by filtering the speech signal through the inverse of the vocal tract model.

The inverse filtering technique has proved to be an efficient method in the estimation of the glottal waveform. If the modeling of the vocal tract can be computed accurately, the result is a fairly good estimate for the real source of the speech production mechanism. Most of the inverse filtering techniques use as their only input the acoustical speech signal. Hence, the analysis is easily and inexpensively arranged.

Even though inverse filtering has become a widely applied analysis procedure it has some drawbacks. First, quite often only a certain kind of speech material can be accurately analysed with an inverse filtering algorithm. The closed phase covariance method, for example, gives reliable results only in the case when the glottal source has a sufficiently long closed phase (Wong et al., 1979). Second, if the identification of the vocal tract is adjusted manually, the estimation of the glottal source is dependent on the subjective criteria applied by the researcher. Finally, the result of inverse filtering is highly dependent on the quality of the input signal. Therefore, great care has to be devoted both to the amplitude and to the phase response of the recording equipment.

In this paper a new glottal wave analysis method, Pitch Synchronous Iterative Adaptive Inverse Filtering (PSIAIF), is presented. The new algorithm is a sequel to two previously presented methods, Adaptive Inverse Filtering (AIF) and Iterative Adaptive Inverse Filtering (IAIF) (Alku et al., 1990, 1991). The structure of the algorithm is presented in Section 2. The performance of the method in the analysis of the glottal excitation is reported in Section 3 using synthetic speech. Section 4 shows the results that were obtained when natural speech was analysed. The author wishes to point out that the main purpose of this paper is to present the new algorithm. A glottal wave study where the emphasis is placed on the detailed analysis of the resulting waveforms is to be published later.

2. Method

2.1. Model

The PSIAIF-method is based on a speech production model that consists of three separate processes: the glottal excitation, the vocal tract and the lip radiation effect (Figure 1). The model is considered to be linear and time-invariant during a certain time interval. The interaction between the different processes is considered to be negligible. Figure 2 shows in the case of a synthetic vowel typical spectral shapes of these processes.

The purpose of the PSIAIF-method is to compute the first part of Figure 1, the glottal excitation, from the speech signal. The last of the three processes, the lip radiation effect, is modeled with a fixed differentiator. Hence, the procedure to compute an accurate estimate for the glottal pulseform concentrates on the identification of the second part of Figure 1, the vocal tract.

2.2. Structure of the PSIAIF-algorithm

2.2.1. General

The PSIAIF-method is a further developed version of the IAIF-method (Iterative Adaptive

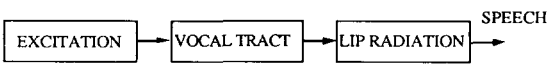


Fig. 1. Separated speech production model.

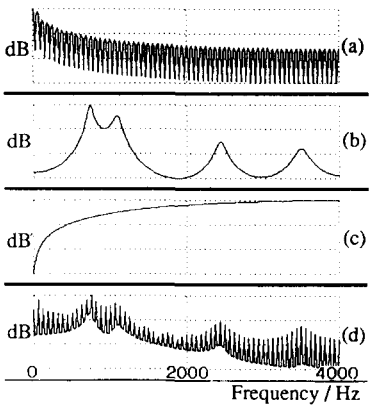


Fig. 2. (a) Spectrum of the glottal excitation. (b) Transfer function of the vocal tract. (c) Transfer function corresponding to the lip radiation effect. (d) Speech spectrum.

Inverse Filtering), that has proved to be promising in the estimation of the glottal excitation (Alku et al., 1991). The IAIF-algorithm works pitch asynchronously. The idea of the IAIF-method, that plays an important role in the PSIAIF-method, is described in Section 2.2.2. The performance of the IAIF-technique is improved in the PSIAIF-algorithm by applying IAIF-analysis twice to the same signal. The first application is pitch asynchronous giving as a result a glottal wave estimate that spans over several pitch periods. This glottal pulseform is used in order to compute an estimate for the length of one glottal cycle. The final estimate for the glottal pulseform is obtained by applying the IAIF-method to speech, over a time interval between two consecutive maximal glottal openings.

2.2.2. *IAIF-method*

The IAIF-method is based on a priori knowledge about the overall shape of the transfer function of the vocal tract (Figure 2(b)). In the case of vowels this is of an all-pass nature having some high energy regions, the formants. If the tilting effect of the glottal source (curve 2(a)) is eliminated from the speech spectrum, the vocal tract can be estimated fairly accurately with linear predictive analysis (LPC). Estimation of the glottal contribution and the vocal tract transfer function is computed in the IAIF-algorithm with an iterative structure that is repeated twice.

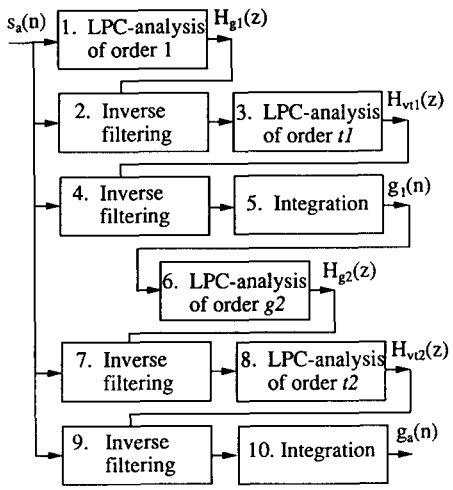
In the beginning of the first iteration the effect of the glottal source to the speech spectrum is modeled by computing an LPC-analysis of order one to the speech signal. By referring to Figure 2 this implies that curve 2(a) is estimated by computing a crude envelope for curve 2(d). The estimated glottal contribution is then cancelled by inverse filtering. A preliminary model for the vocal tract is obtained by applying higher order LPC-analysis to the signal from which the effect of the source was cancelled. The first estimate for the glottal excitation is obtained by cancelling from the original speech signal the effects of the tract and lip radiation by inverse filtering.

The resulting glottal wave estimate of the first iteration is used in the beginning of the second iteration in order to compute a more accurate model for the glottal contribution. The effect of the glottal excitation to the speech spectrum is estimated in the second iteration using an LPC-analysis whose order equals, typically, two or four. It is important to note that the estimation of the glottal contribution during *the first iteration* shall not be computed using an LPC-order that is higher than one. A consequence of using higher order LPC-analysis would be to model the formants by a filter which models the effect of the source.

After cancelling the estimated glottal contribution the model for the vocal tract is formed once again using higher order LPC-analysis. The final result is obtained by inverse filtering the effects of the vocal tract and lip radiation from the original speech signal.

The block diagram of the IAIF-method is shown in Figure 3. The speech signal to be analysed is denoted $s_a(n)$ and the result, the pitch asynchronously computed estimate for the glottal excitation, is denoted $g_a(n)$. The first iteration consists of the blocks numbered from 1 to 5 and the second iteration of the blocks numbered from 6 to 10. The purpose of each of the blocks is described as follows.

Block no. 1:
The effect of the glottal pulseform to the speech spectrum is preliminarily estimated by first order LPC-analysis.
Block no. 2:
The estimated glottal contribution is eliminated by filtering $s_a(n)$ through $H_{g1}(z)$.



Where:

$H_{g1}(z) = 1 + az^{-1}$ $\quad H_{g2}(z) = 1 + \sum_{k=1}^{g2} c(k)z^{-k}$

$H_{vt1}(z) = 1 + \sum_{k=1}^{t1} b(k)z^{-k}$ $\quad H_{vt2}(z) = 1 + \sum_{k=1}^{t2} d(k)z^{-k}$

Fig. 3. Structure of the IAIF-algorithm.

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 $s_a(n)$ 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 starts by computing a new estimate for the effect of the source to the speech spectrum. This time LPC-analysis of order $g2$ 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 $t2$ to the output of the previous block.
Block no. 9:
The effect of the vocal tract is eliminated from speech by filtering $s_a(n)$ through $H_{vt2}(z)$.
Block no. 10:
The result, $g_a(n)$, is obtained by cancelling the lip radiation effect by integrating the output of Block no. 9.

2.2.3. *PSIAIF-method*

The IAIF-algorithm that was described in the previous section computes the glottal excitation pitch asynchronously over several fundamental periods. In the estimation of the vocal tract transfer function LPC-analysis has an important role. However, it is known (Makhoul, 1975) that formant estimation with linear predictive analysis can be seriously affected by the harmonic structure of the speech spectrum. It might happen that instead of modeling a formant, LPC-analysis models a spectral peak that results from the periodical pulseform of the excitation.

In order to improve the performance of LPC-analysis in the estimation of the vocal tract transfer function the final glottal wave estimate is computed pitch synchronously. Hence, the speech spectrum is free from the harmonic structure of the source. Another advantage of using pitch synchronous computation is more accurate modeling of the formants in the case when the vowel rapidly changes from one to another.

It should be noted that pitch synchronous analysis does not completely remove the drawbacks of linear predictive analysis in the estimation of the vocal tract transfer function. An accurate analysis of speech whose fundamental frequency is high and first formant is low is a difficult task. In the time domain this implies that in female voices, for example, the tract is excited by a source pulseform in a way that the effect of the previous pulse will not be entirely attenuated before the next one occurs. As a consequence the performance of LPC-analysis in the estimation of the formants will decrease (Makhoul, 1975).

The flow diagram of the PSIAIF-algorithm is shown in Figure 4. The speech signal to be analysed is denoted $s(n)$ and the final result, the estimate for the glottal excitation, is denoted $g(n)$. In the beginning of the algorithm the speech signal $s(n)$ is high-pass filtered in order to remove undesirable fluctuations of the result. The high-pass filter is a linear phase FIR whose cut-off frequency is 30 Hz.

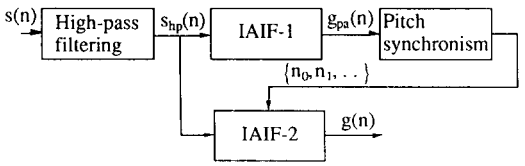


Fig. 4. Structure of the PSIAIF-algorithm.

The high-pass filtered speech signal, $s_{hp}(n)$, is used as an input to the first IAIF-analysis (block "IAIF-1" in Figure 4). As a result, one frame of a pitch asynchronously computed glottal wave estimate, $g_{pa}(n)$, is obtained. In the next stage of the algorithm parameters of the pitch synchronism are estimated from $g_{pa}(n)$. The length of the pitch period, denoted M, is first computed by searching for the maximum of the autocorrelation of $g_{pa}(n)$ between time indices 20 and 120 (corresponding to $F0$ equal to 400 and 67 Hz, respectively, when the sampling frequency is 8 kHz). The time indices of maximum glottal openings, $\{n_0, n_1, \ldots, n_R\}$, are computed for the frame of $g_{pa}(n)$ as shown in Figure 5. The time instant of the maximum glottal opening, n_i, is determined by searching for the absolute maximum of $g_{pa}(n)$ within a certain time span. This time span, as shown by the smaller arrow in Figure 5, includes M samples the first of which occurs at index $n_{i-1}+0.5M$.

The final estimate for the glottal excitation is obtained by analysing the high-pass filtered speech signal, $s_{hp}(n)$, with the IAIF-algorithm pitch synchronously (block "IAIF-2" in Figure 4). The IAIF-analysis is computed from signal $s_{hp}(n)$ one fundamental period at a time using frames that span between two consecutive maximal glottal openings.

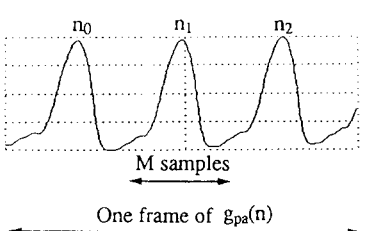


Fig. 5. Determination of time instants of maximal glottal openings from $g_{pa}(n)$. (Search region for n_1 using the index of the previous maximal glottal opening, n_0, and the length of the pitch period, M, is shown by the upper arrow.)

3. Analysis of synthetic vowels

In order to verify the performance of the PSIAIF-method the new algorithm was first tested using synthetic speech. Two vowels ([a] and [i]) were created using a procedure described in (Gold and Rabiner, 1968). Vocal tracts were simulated in the synthesis procedure using eighth order all-pole filters. The bandwidth of the signals was 4 kHz. As a synthetic source we used a glottal wave model suggested in (Ananthapadmanabha, 1984). The shape of the glottal pulseform was changed from one period to another in order to simulate speech production where phonation slides from breathy to pressed. Synthetic speech corresponding to female and male speakers was created for the two vowels. Synthetic female voices were produced by changing the length of the pitch period during sliding of phonation from 60 samples (corresponding to $F0$ equals 133 Hz) to 40 samples (corresponding to $F0$ equals 200 Hz). The same values for the male vowels were 120 samples (corresponding to $F0$ equals 67 Hz) and 80 samples (corresponding to $F0$ equals 100 Hz), respectively.

The synthetic vowels were analysed with the PSIAIF-algorithm that has been implemented on a Symbolics Lisp-machine using the QuickSig signal processing environment (Karjalainen et al., 1988). By referring to Figure 4 the block length of the IAIF-1-analysis was 256 samples (32 ms). Orders of LPC-analysis in Figure 3 were as follows: $t1=10$, $g2=4$ and $t2=10$. Male voices were analysed by using autocorrelation criteria together with Hamming-windowing in each of the LPC-analysis. In female speech the length of the fundamental period is usually quite short. Hence, the number of speech samples that are used in the LPC-analysis of the IAIF-2-stage is fairly small. Therefore covariance criteria was used in the pitch synchronous IAIF-2-stage. If the inverse filter obtained by linear prediction had zeros outside the unit circle, they were replaced by their mirror image partners inside the unit circle.

In order to compare the performance of the PSIAIF-algorithm to other glottal wave analysis methods the synthetic speech material was also analysed using the closed phase covariance technique (Wong et al., 1979). This method is based on the idea that the vocal tract is identified by

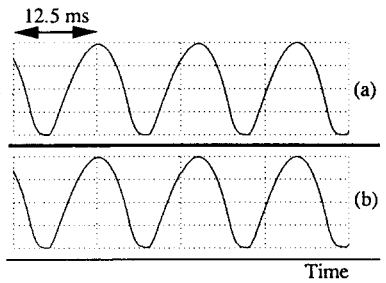


Fig. 6. Glottal waveforms of a synthetic [a]-vowel corresponding to breathy male phonation. (a) Original excitation. (b) Estimate given by the PSIAIF-method.

computing LPC-analysis from those samples during which there is no glottal excitation. The order of the covariance analysis was eight. The length of the covariance frame (i.e., the number of samples during which the prediction error of LPC-analysis is minimized) was 12 samples. The vocal tract filter was updated once every 32 ms. The result given by the closed phase analysis is highly dependent on the position of the covariance frame. Therefore, the analysis was computed for the signals by varying the beginning of the frame. The waveform that was most similar to the original excitation was selected as the result.

When the vowel [a] from a male voice was analysed, the waveform given by the PSIAIF-algorithm was very similar to the original excitation. This holds true during the entire signal without depending on the type of phonation as can be seen in Figures 6 and 7. Only at the end of the signal, where phonation was pressed, a small ripple component could be distinguished during the closed phase of the glottal cycle.

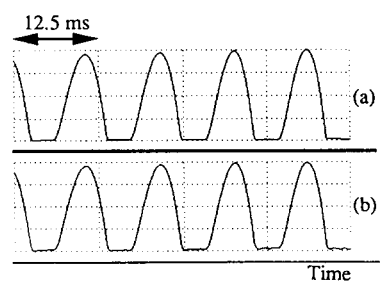


Fig. 7. Glottal waveforms of a synthetic [a]-vowel corresponding to pressed male phonation. (a) Original excitation. (b) Estimate given by the PSIAIF-method.

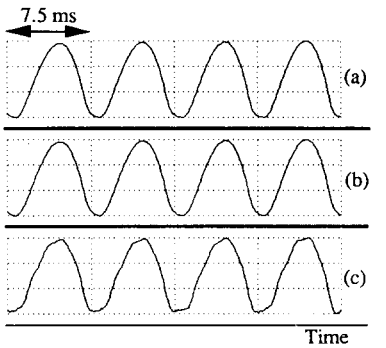


Fig. 8. Glottal waveforms corresponding to a synthetic female [a]-vowel of breathy phonation. (a) Original excitation. (b) Estimate given by the PSIAIF-method. (c) Estimate given by the closed phase covariance method.

In the case of the female [a]-vowel, the quality of the results decreased only slightly compared to the analysis of the male vowel. As can be seen from Figure 8 the PSIAIF-algorithm yielded a pulseform that was very similar to the original source during breathy phonation. In the case of pressed phonation (Figure 9), the result was also very close to the original source signal. A typical distortion in the case of pressed phonation was a slight smoothing of the pulseform at the moment of glottal closure.

When the performance of the PSIAIF-algorithm was compared to the closed phase covariance technique, certain differences were found. The results given by the closed phase analysis were accurate only when signals of sufficiently long closed phases were analysed provided that the analysis frame was at the correct position. The problems of the closed phase covariance procedure can be clearly seen in the analysis of female speech of breathy phonation.

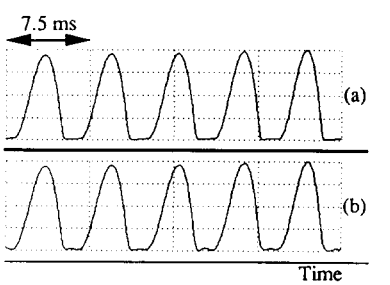


Fig. 9. Glottal waveforms corresponding to a synthetic female [a]-vowel of pressed phonation. (a) Original excitation. (b) Estimate given by the PSIAIF-method.

The reason is that in the case of breathy phonation the glottal pulseform does not have a clear closed phase. Hence, the vocal tract is estimated poorly by the covariance analysis since there is no time interval during which the assumption of zero excitation would be correct. A typical result is shown by curve 8(c). It is interesting to note that the closed phase analysis very often forces the resulting pulseform to comprise a kind of closed phase even though the original excitation does not have a clear closed phase.

When synthetic [i]-vowels were analysed, the results obtained by the PSIAIF-method were very close to the original excitation during breathy and normal phonation. When the vowel [i] was analysed during pressed phonation the result was distorted by a ripple component. Distortion was largest in the case of the female voice. The reason for the deterioration of the results in the analysis of pressed female [i]-vowel is obvious. The vowel [i] has a low first formant. The spectrum of the source signal decays more slowly in pressed phonation than in breathy or normal phonation. Hence, the adaptive prefiltering (Block no. 6 of Figure 3) does not eliminate the glottal contribution properly enough. As a result, the first formant is estimated poorly by the vocal tract LPC-analysis (Block no. 8 of Figure 3) and a formant ripple will be present in the final result.

4. Analysis of natural vowels

4.1. Speech material

The PSIAIF-algorithm was used in the estimation of the glottal excitation of natural speech by studying vowels of four female and four male speakers. Each of the subjects were of healthy voice. Two experiments were performed. In the first one, the subjects were asked to produce the vowel [a] using breathy, normal and pressed phonation. With this material our target was to find out how the new algorithm was able to analyse different phonation types from natural speech. In the second experiment the subjects were asked to produce short words of the Finnish language using their natural phonation type. Each of the words ("ou" [ou], "ai" [ai], "ui" [ui] and "ei" [ei]) consisted of two different vowels. The target of the second experiment was to study how the pitch synchronous algorithm was able to analyse the glottal excitation in the case when the vocal tract was continuously changing.

The speech signals were recorded in an anechoic chamber using a condenser microphone (Brüel&Kjær 4134). The data was AD-converted using a Sony PCM-F1 and collected on a video cassette with a Sony SL-F1E tape recorder. The bandwidth of the signals was reduced in the Symbolics-computer from the original value of 20 kHz to 4 kHz.

4.2. Results

All the signals were analysed with the PSIAIF-method using the following parameters (Figure 3): $t1=12$, $g2=4$, $t2=12$. The block length of the pitch asynchronous IAIF-analysis (block IAIF-1 in Figure 4) was 256 samples (32 ms). Autocorrelation criteria together with Hamming-windowing was used in all of the LPC-analysis of the IAIF-1-stage. Male voices were analysed in the pitch synchronous IAIF-2-stage using autocorrelation criteria together with Hamming-windowing. The IAIF-2-stage for female voices were computed using covariance criteria.

EXPERIMENT 1. Figure 10 shows typical waveforms that were obtained when the vowel [a] of male speakers was analysed. Curves 10(a), 10(b) and 10(c) correspond to the analysis of breathy, normal and pressed phonation, respectively. The

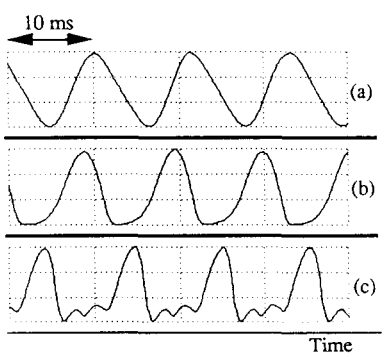


Fig. 10. Glottal waveforms corresponding to a natural male [a]-vowel. (a) Breathy phonation. (b) Normal phonation. (c) Pressed phonation.

obtained waveforms are of quite reliable shapes. The glottal wave of breathy phonation is smooth. The maximal glottal opening occurs approximately in the middle of the glottal cycle. The pulseform shows that right after the time instant of the minimal flow the vocal folds start to open, i.e. there is no specific closed phase during the glottal cycle. In the case of normal phonation (curve 10(b)) the glottal waveform has a nearly flat closed phase. The time instant of the maximal glottal opening has moved towards the end of the cycle. The pulseform corresponding to pressed phonation (curve 10(c)) was partly distorted by a formant ripple. The author wishes to emphasize that the speech signal from which curve 10(c) was computed was produced using a very pressed phonation type. Hence, the slowly decaying spectrum of the source deteriorates the accurate estimation of the first formant.

Figure 11 shows the results that were obtained when female [a]-vowels of three different phonation types were analysed. The pulseform corresponding to breathy phonation (Figure 11(a)) is quite similar to the one that was obtained from the male voice (curve 10(a)). The shape of the glottal pulseform did not change much in the case of the female voice when phonation changed from breathy to normal. Only the time instant of the maximal glottal opening seemed to move slightly towards the end of the glottal cycle. However, in the case of pressed phonation, as can be seen from curve 11(c), the shape of the glottal excitation changed radically. The pulseform of curve 11(c) is characterized by a clear closed phase. Both the opening and closure of the glottis occurred much more abruptly than in the case of normal phonation.

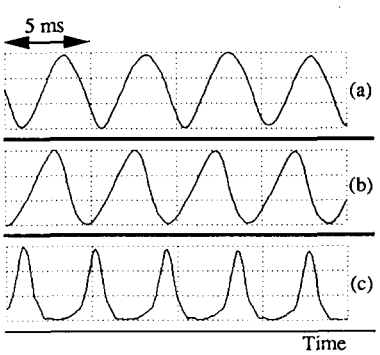


Fig. 11. Glottal waveforms corresponding to a natural female [a]-vowel. (a) Breathy phonation. (b) Normal phonation. (c) Pressed phonation.

EXPERIMENT 2. When the short words produced by male speakers were analysed, it was found that for two of the subjects the shape of the obtained pulseform remained almost the same independent of the vowel that was analysed. Figure 12 shows typical waveforms. This result is well in line with an assumption that the shape of the glottal source does not depend much on the vowel that is produced. However, for the other two subjects the waveform changed when the vowel [i] was analysed. A typical behaviour of the estimated glottal source for these two subjects was that in the word "ai", for example, the resulting glottal pulseform had a closed phase during the vowel [a]. When the analysis was computed during the vowel [i], the closed phase disappeared. The reason for this behaviour might be that the estimation of the vocal tract decreased because the low first formant could not be satisfactorily separated from the source. Another explanation is that the subjects really changed their glottal pulseform during the production of the voices. In order to know which one of these two explanations is correct it would have been of great help to use other techniques (laryngograph, for example) together with inverse filtering analysis. Unfortunately, the laryngograph apparatus was not available when the recordings were made.

A comparison was also made between the result given by the pitch asynchronous IAIF-algorithm (signal $g_{pa}(n)$ of Figure 4) and the pitch synchronous PSIAIF-method (signal $g(n)$ of Figure 4). It was found that the pitch synchronous analysis improved the result. Especially for vowels with a

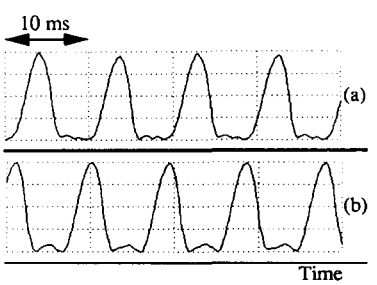


Fig. 12. Glottal waveforms corresponding to a natural male voice. (a) Vowel [a] in word "ai". (b) Vowel [i] in word "ai".

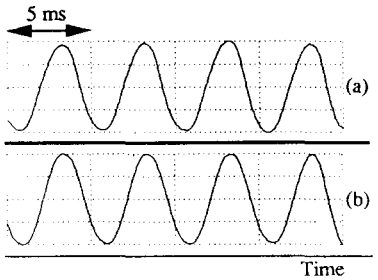


Fig. 13. Glottal waveforms corresponding to a natural female voice. (a) Vowel [u] in word "ui". (b) Vowel [i] in word "ui".

low first formant ([i] and [u]) the ripple component during the closed phase was clearly smaller when the analysis was computed pitch synchronously. The improvements are explained by the fact that in pitch asynchronous analysis the harmonic structure of the speech spectrum deteriorates the estimation of the vocal tract by LPC-analysis.

When words produced by the four female subjects were analysed, it was found that the shapes of the resulting glottal pulseforms remained quite constant between different vowels. Figure 13 shows a typical result. In general the excitation waveforms were of smooth shapes and the lengths of the closed phases were short. The reason why the results were not very much dependent on the vowel is explained by a spectrum of a typical female glottal pulse. Since the source spectrum decays quite quickly, the estimation of the first formant can be obtained fairly accurately even for vowels [i] and [u].

5. Conclusions

In this paper a new glottal wave analysis method, Pitch Synchronous Iterative Adaptive Inverse Filtering (PSIAIF), was presented. The algorithm is based on Iterative Adaptive Inverse Filtering (IAIF). In the IAIF-method the average contribution of the glottal excitation to the speech spectrum is estimated with an iterative structure. After eliminating the estimated glottal contribution the vocal tract is identified using LPC-analysis. Finally, the estimate for the glottal pulseform is obtained by eliminating the effects of the vocal tract and lip radiation by inverse filtering. In the new PSIAIF-algorithm the speech signal is first analysed pitch asynchronously by the IAIF-method. The obtained glottal pulseform is used in order to compute time instants of maximal glottal openings. A final estimate for the glottal source is obtained by processing the original speech signal by the IAIF-algorithm pitch synchronously using speech samples between consecutive maximal glottal openings.

By analysing synthetic speech the new method was shown to give a fairly good estimate for the glottal excitation. This holds for different kinds of phonation types. Female and male voices can be analysed with nearly equal accuracy. The only case when the result was distorted was the analysis of the synthetic female [i]-vowel that was created with a very pressed phonation type.

The PSIAIF-method was used in the analysis of natural speech by studying the vowel [a] of three different phonation types (breathy, normal and pressed) and short words that consisted of different vowels. The glottal excitations given by the new algorithm were of reliable shapes for most of the signals. The obtained glottal flows were partly distorted by a formant ripple in the analysis of vowels with very pressed phonation type. For male voices of normal phonation type, the analysis of the vowel [i] sometimes gave results that were distorted because of the poor estimation of the first formant.

The author wishes to point out some features of the PSIAIF-method that justify its use in glottal wave analysis. First, as was shown in this paper, the algorithm is able to give a fairly good estimate for the glottal excitation from different speech signals. Second, the algorithm is fully automatic. Hence, it is easy to use and the results do not depend on the subjective criteria applied by the investigator. Third, the method is non-invasive. Finally, the computational complexity of the algorithm is not very high. Hence, it is possible to implement the algorithm in real-time applications.

Acknowledgment

The author wishes to express his gratitude to Dr. Unto K. Laine, Prof. Matti Karjalainen and Prof. Erkki Vilkman for their comments. A warm thanks also to Toomas Altosaar and Petra Wingert for their help.

References

P. Alku, E. Vilkman and U.K. Laine (1990), "A comparison of EGG and a new inverse filtering method in phonation change from breathy to normal", *Proc. 1st Internat. Conf. Spoken Language Processing*, Kobe, 18–22 November 1990, pp. 197–200.

P. Alku, E. Vilkman and U.K. Laine (1991), "Analysis of glottal waveform in different phonation types using the new IAIF-method", *Proc. 12th Internat. Congress Phonetic Sciences*, Vol. 4, Aix-en-Provence, 19–24 August 1991, pp. 362–365.

T.V. Ananthapadmanabha (1984), "Acoustic analysis of voice source dynamics", Speech Transmission Laboratory Quarterly Progress and Status Report, Nos. 2–3, Royal Institute of Technology, Stockholm, Sweden, pp. 1–24.

B. Gold and L.R. Rabiner (1968), "Analysis of digital and analog formant synthesizers", *IEEE Trans. Audio Electroacoust.*, Vol. 16, pp. 81–94.

M.J. Hunt, J.S. Bridle and J.N. Holmes (1978), "Interactive digital inverse filtering and its relation to linear prediction methods", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.*, pp. 15–18.

M. Karjalainen, T. Altosaar and P. Alku (1988), "QuickSig - An object-oriented signal processing environment", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.*, New York, 11–14 April 1988, pp. 1682–1685.

A.K. Krishnamurthy and D.G. Childers (1986), "Two-channel speech analysis", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 34, pp. 730–743.

J. Makhoul (1975), "Linear Prediction: A tutorial review", *Proc. IEEE*, Vol. 63, pp. 561–580.

J.D. Markel and A.H. Gray, Jr. (1976), *Linear Prediction of Speech* (Springer, New York).

D.Y. Wong, J.D. Markel and A.H. Gray, Jr. (1979), "Least squares glottal inverse filtering from the acoustic speech waveform", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 27, pp. 350–355.