# A Review of Glottal Waveform Analysis

Jacqueline Walker and Peter Murphy

Department of Electronic and Computer Engineering,
University of Limerick,
Limerick, Ireland
jacqueline.walker@ul.ie,peter.murphy@ul.ie

## 1 Introduction

Glottal inverse filtering is of potential use in a wide range of speech processing applications. As the process of voice production is, to a first order approximation, a source-filter process, then obtaining source and filter components provides for a flexible representation of the speech signal for use in processing applications. In certain applications the desire for accurate inverse filtering is more immediately obvious, e.g., in the assessment of laryngeal aspects of voice quality and for correlations between acoustics and vocal fold dynamics, the resonances of the vocal tract should firstly be removed. Similarly, for assessment of vocal performance, trained singers may wish to obtain quantitative data or feedback regarding their voice at the level of the larynx.

In applications where the extracted glottal signal is not of primary interest in itself the goal of accurate glottal inverse filtering remains important. In a number of speech processing applications a flexible representation of the speech signal, e.g., harmonics plus noise modelling (HNM) [74] or sinusoidal modelling [65], is required to allow for efficient modification of the signal for speech enhancement, voice conversion or speech synthesis. In connected speech it is the glottal source (including the fundamental frequency) that changes under a time-varying vocal tract and hence an optimum representation should track glottal and filter changes. Another potential application of glottal inverse filtering is speech coding, either in a representation similar to HNM, for example (but incorporating a glottal source), or as in [3], [11], [19] applying coding strategies termed glottal excited linear prediction (GELP) which use glottal flow waveforms to replace the residual or random waveforms used in existing code excited linear prediction (CELP) codecs. In the studies cited the perceptual quality of the GELP codecs is similar to that of CELP.

The speaker identification characteristics of glottal parameters have also recently undergone preliminary investigation [64] (note this is quite different from investigating the speaker identification characteristics of the linear prediction residual signal). Identification accuracy up to approximately 70% is reported using glottal parameters alone. Future studies employing explicit combinations of glottal and filter components may provide much higher identification rates. In addition, the more that is understood regarding glottal changes in connected speech

Y. Stylianou, M. Faundez-Zanuy, A. Esposito (Eds.): WNSP 2005, LNCS 4391, pp. 1–21, 2007.

the better this knowledge can be used for speaker identification (or conversely it may lead to improved glottal de-emphasis strategies for speech recognition).

Due to non-availability of a standard, automatic GIF algorithm for use on connected speech (perhaps due to a lack of knowledge of the overall voice production process), representation and processing of the speech signal has generally side stepped the issue of accurate glottal inverse filtering and pragmatic alternatives have been implemented. However these alternatives come at a cost, which can necessitate the recording of considerably more data than would be required if the dynamics of voice production were better understood and the appropriate parameters could be extracted. For example, in existing methods for pitch modification of voiced speech, the glottal parameters are not manipulated explicitly, e.g., in the sinusoidal model a deconvolution is performed to extract the filter and residual error signal. The linear prediction residual signal (or the corresponding harmonic structure in the spectrum) is then altered to implement the desired pitch modification. This zero-order deconvolution ensures that the formant frequencies remain unaltered during the modification process, giving rise to a shape-invariant pitch modification. However, examining this from a voice production viewpoint reveals two important consequences of this approach: to a first approximation, the glottal closed phase changes and eventually overlaps as the fundamental frequency (f0) increases and if the glottal periods are scaled the spectral tilt will change [38]. The former may explain the hoarseness reported in [65] after 20% modification. The solution to this problem has been to record more data over a greater range of f0 and always modify within this limit. Integrating better production knowledge into the system would facilitate modification strategies over a broader range without recourse to such an extensive data set. In what follows, we review the state of the art in glottal inverse filtering and present a discussion of some of the important issues which have not always been at the forefront of consideration by investigators. After first establishing a framework for glottal waveform inverse filtering, the range of approaches taken by different investigators is reviewed. The discussion begins with analog inverse filtering using electrical networks [53] and extends to the most recent approaches which use nonlinear least squares estimation [64]. A brief review of the earliest approaches shows that most of the basic characteristics of the glottal waveform and its spectrum were established very early. There was also interest in developing specialist equipment which could aid recovery of the waveform. With the introduction of the technique of linear prediction and the steady improvement of computing power, digital signal processing techniques came to dominate. Although parametric modelling approaches have been very successful, alternatives to second-order statistics have not been used extensively and have not, so far, proved very productive. As the glottal waveform is a low frequency signal, recording conditions and phase response play a very important role in its reconstruction. However, sufficient attention has not always been paid to these issues. Finally, the question of identifying a good result in the reproduction of such an elusive signal is discussed.

## 2 A Framework for Glottal Waveform Inverse Filtering

Linear prediction is a very powerful modelling technique which may be applied to time series data. In particular, the all-pole model is extensively used. In this model, as shown in (1), the signal is represented as a linear combination of past values of the signal plus some input [47]:

$$s_n = \sum_{k=1}^{p} a_k s_{n-k} + Ap\,(n) \ . \tag{1}$$

where $A$ is a gain factor applied to the input $p\,(n)$. In the frequency domain, this model represents an all-pole filter applied to the input:

$$V\,(z) = \frac{A}{1 + \sum_{k=1}^{p} a_k z^{-k}} \ . \tag{2}$$

As is well known, using the method of least squares, this model has been successfully applied to a wide range of signals: deterministic, random, stationary and non-stationary, including speech, where the method has been applied assuming local stationarity. The linear prediction approach has been dominant in speech due to its advantages:

1. Mathematical tractability of the error measure (least squares) used.
2. Favorable computational characteristics of the resulting formulations.
3. Wide applicability to a range of signal types.
4. Generation of a whitening filter which admits of two distinct and useful standard input types.
5. Stability of the model.
6. Spectral estimation properties.

Applied to the acoustic wave signal, linear prediction is used to produce an all-pole model of the system filter, $V\,(z)$, which turns out to be a model of the vocal tract and its resonances or formants. As noted above, the assumed input to such a model is either an impulse or white noise, both of which turn out to suit speech very well. White noise is a suitable model for the input to the vocal tract filter in unvoiced speech and an impulse (made periodic or pseudo-periodic by application in successive pitch periods) is a suitable model for the periodic excitation in voiced speech. In the simplest model of voiced speech, shown in Fig. 1, the input is the flow of air provided by the periodic opening and closing of the glottis, represented here by a periodic impulse train $p\,(n)$. The vocal tract acts as a linear filter, $v\,(n)$, resonating at specific frequencies known as formants. Speech, $s\,(n)$, is produced following radiation at the lips represented by a simple differentiation, $r\,(n)$.

### 2.1 Closed Phase Inverse Filtering

In linear prediction, the input is assumed unknown: the most information we can recover about the input is a prediction of its equivalent energy [47]. As

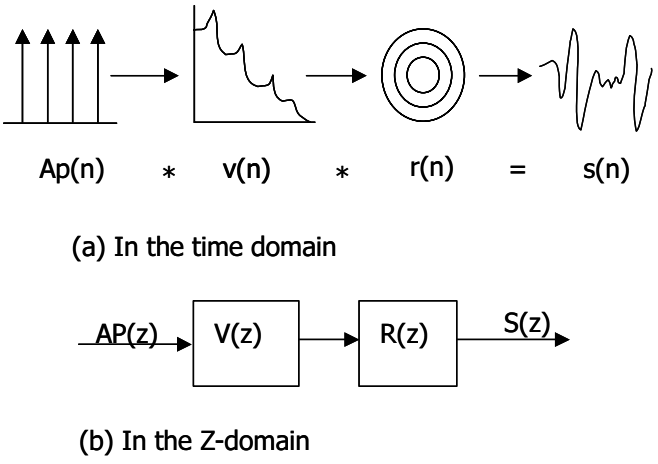



**Fig. 1.** Simplest model of voiced speech

a consequence of the least squares modelling approach, two models fit the assumptions of linear prediction: the input impulse and white noise. Both of these inputs have a flat spectrum. In other words, the inverse filter which results from the process is a whitening filter and what remains following inverse filtering is the modelling error or residual. The simplest glottal pulse model is the periodic impulse train [21] as used in the LPC vocoder [78]. However, speech synthesizers and very low bit rate speech coders using only periodic impulses and white noise as excitations have been found to be poor at producing natural sounding speech. To improve speech quality, it has been found useful to code the residual, for example using vector quantized codebooks, in speech coding techniques such as CELP [68], since to the extent that the residual differs from a purely random signal in practice, it retains information about the speech including the glottal waveform.

The linear speech production model can be extended as shown in Fig. 2 so that it includes two linearly separable filters [21]. The glottal excitation, $p(n)$ does not represent a physical signal but is simply the mathematical input to a filter which will generate the glottal flow waveform, $g(n)$. In this model, lip radiation is represented by a simple differencing filter:

$$R(z) = 1 - z^{-1} \ . \tag{3}$$

and glottal inverse filtering requires solving the equation:

$$P(z)\,G(z) = \frac{S(z)}{AV(z)\,R(z)} \ . \tag{4}$$

To remove the radiation term, define the differentiated glottal flow waveform as the effective driving function:

$$Q(z) = P(z)\,G(z)\,R(z) \ . \tag{5}$$

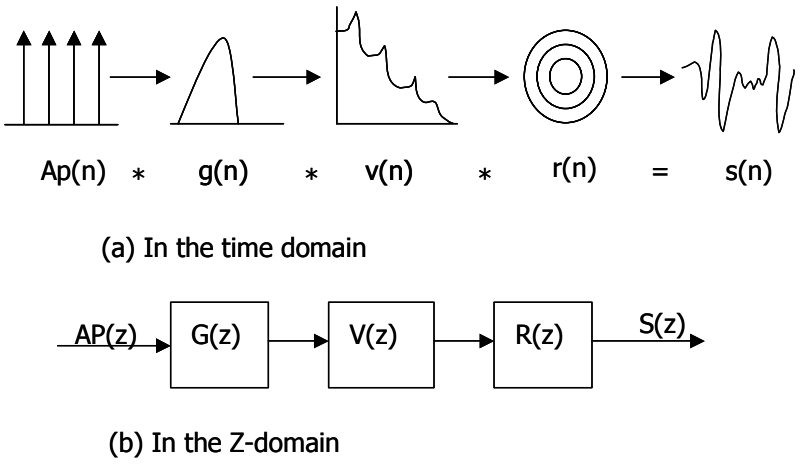


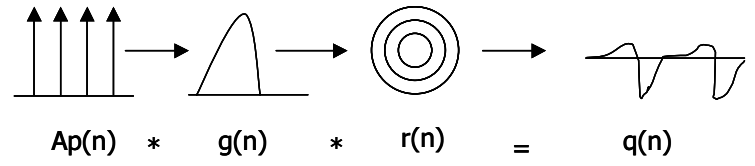

**Fig. 2.** The linear speech model with linearly separable source model

as shown in part (a) of Fig. 3. Now, as shown in part (b) of Fig. 3, inverse filtering simplifies to:

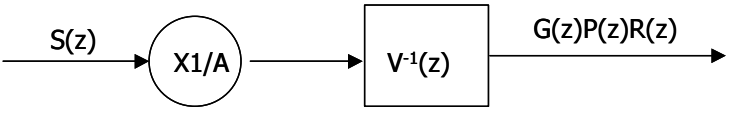
$$Q(z) = \frac{S(z)}{AV(z)} \quad . \tag{6}$$

To solve for both $Q(z)$ and $V(z)$ is a blind deconvolution problem. However, during each period of voiced speech the glottis closes, $g(n) = 0$, providing an impulse to the vocal tract. While the glottis is closed, the speech waveform must be simply a decaying oscillation which is only a function of the vocal tract and its resonances or formants [81] i.e. it represents the impulse response of the vocal tract. Solving for the system during this closed phase should exactly capture the vocal tract filter, $V(z)$, which may then be used to inverse filter and recover $Q(z)$. $G(z)$ may then be reconstructed by integration (equivalently by inverse filtering by $\frac{1}{R(z)}$). This approach is known as closed phase inverse filtering and is the basis of most approaches to recovering the glottal flow waveform.

**Early Analog Inverse Filtering.** According to [81], in analog inverse filtering, a "physically meaningful mathematical basis for glottal inverse filtering has not been explicitly applied."(p. 350), rather the glottal closed region was estimated and parameters were adjusted until a smooth enough glottal waveform emerged. This description is perhaps a bit unfair as the first inverse vocal tract filters were analog networks as in [53] which had to be built from discrete components and required laborious tuning. Because of these difficulties, in [53] only the first two formants were considered and removed, which led to considerable ripple in the closed phase of the recovered glottal waveforms. Nevertheless the recovered glottal waveforms were quite recognizable. A ripple on the glottal waveform corresponding to the first formant was also noted a sign of inaccurate first formant




(a) In the time domain




(b) Applied to inverse filtering in the Z-domain

**Fig. 3.** The effective driving function

estimation and a capacitor in the inverse filter network could be adjusted until the ripple disappeared. As well as attempting to recover the glottal waveform in the time domain, it could be modelled in the frequency domain. In [49], a digital computer was used to perform a pitch synchronous analysis using successive approximations to find the poles and zeros present in the speech spectrum. With the assumption of complete glottal closure, there will be discontinuous first derivatives at the endpoints of the open phase of the glottal waveform, 0 and $T_c$, and a smooth glottal waveform open phase (i.e. the second derivative exists and is bounded), it is then shown that the glottal waveform can be modelled by a set of approximately equally spaced zeros $\sigma + j\omega$ where:

$$\sigma = \ln(\frac{-g\prime(T_c-)}{g\prime(0+)}) \tag{7}$$

$$\omega = \frac{\pi}{T_c}(1 \pm 2n) \tag{8}$$

Thus, the vocal tract is modelled by poles and the glottal waveform by zeros. Despite its simplicity, this approach to finding a model of the glottal waveform has not often been pursued since. Most pole-zero modelling techniques are applied to find vocal tract zeros, especially those occurring in nasal or consonant sounds. Furthermore, it is often pointed out that it is not possible to unambiguously determine whether zeros 'belong' to the glottal waveform or the vocal tract. However, there are exceptions [48].

Glottal waveform identification and inverse filtering with the aim of developing a suitable glottal waveform analog to be used in speech synthesis was first attempted in [66]. In this work, inverse filtering was performed pitch synchronously over the whole pitch period. A variety of waveforms ranging from

simple shapes such as a triangle and a trapezoid through to piecewise sinusoids which more closely matched the inverse filtered glottal waveform shape were used to synthesize speech which was assessed for naturalness in listening tests. It was found that the shapes which produced the most natural sounding speech had a spectral decay of 12 dB/octave: consistent with continuous functions with discontinuous first and higher order derivatives, as postulated by [49].

**Approaches Requiring Special Equipment.** The Rothenberg mask was introduced to overcome some of the perceived difficulties with inverse filtering, in particular: susceptibility to low frequency ambient noise, the difficulty of amplitude calibration and the inability to recover the correct DC offset level [67]. The mask is a specially vented pneumotachograph mask which permits direct measurement of the oral volume velocity and so, by eliminating the lip radiation component, removes the pole at DC in the inverse filter. According to [67], knowledge of the glottal waveform down to DC allows for periods of glottal closure to be identified absolutely, rather than by the relative flatness of the wave. The main disadvantage of the mask is its limited frequency response which extends to only 1 kHz and limits its ability to reasonably resolve glottal waveforms to those with a fundamental frequency of at most about 200 Hz [67]. While not limiting its applicability entirely to the speech of adult males, this does make its range "somewhat inadequate for most female speakers and for children."([67], p. 1637). However, the Rothenberg mask has been applied to study glottal waveforms in female speakers [33]. While useful for precise laboratory-based studies, inverse filtering using the Rothenberg mask is obviously not suited to many other applications such as speech coding and speaker identification as it requires special equipment, trained operators and takes some time to apply.

A second equipment based approach was introduced by Sondhi [70] who pointed out that speaking into a reflectionless, rigid-walled tube can cancel out the vocal tract contribution allowing the investigator simply to record the glottal waveform directly by a microphone inserted into the wall of the tube. Here, the equipment is relatively cheap and easy to construct and, as for the Rothenberg mask, provides some built-in protection against low frequency ambient noise. A number of subsequent investigations have employed this technique [56], [57], [71]. Compared with the waveforms recovered using the Rothenberg mask, the waveforms recovered using the Sondhi tube have a certain peakiness. A detailed study of the frequency response of the tube set-up in [56] showed that while the tube was effective down to frequencies of 90 Hz, it had a resonance at 50 Hz and the low frequency response (below 50 Hz) had 'additional factors' (quite what these were is not made fully clear, but it could have been simply noise) which were removed by high pass filtering by the microphone and pre-amplifier (at 20 Hz and 30 Hz respectively). Compensation for the resulting high pass characteristic of the overall system removed peakiness from the glottal waveforms. As Sondhi also applied high pass filtering with a cut-off of 20 Hz and did not compensate for it, this could be the cause of the observed peaky shape of the recovered waveforms. Another possible source of distortion is the acoustic load

or impedance provided by the tube which could have been too high [32]. Mask (or tube) loading can cause attenuation of and, more importantly, shifts in the formants [67].

**Approaches Requiring a Second Channel.** The main difficulty in closed phase inverse filtering is to identify precisely the instants of glottal closure and opening. For example, in [81], the closed phase is identified as that part of the waveform for which the normalized total squared error is below some threshold. As it is more forceful, glottal closure may be identified more easily than glottal opening, which is more gradual [81]. Due to these difficulties, some investigators have made use of the electroglottography (EGG) signal to locate the instants of glottal closure and opening [17], [43],[44],[51],[79]. In particular, it is claimed that use of the EGG can better identify the closed phase in cases when the duration of the closed phase is very short as in higher fundamental frequency speech (females, children) or breathy speech [79]. As with the methods requiring special equipment, two-channel methods are not useful for more portable applications or those requiring minimal operator intervention. However, precisely because they can identify the glottal closure more accurately, results obtained using the EGG can potentially serve as 'benchmarks' by which other approaches working with the acoustic pressure wave alone can be evaluated. The same is clearly true of the equipment-based approaches as long as the characteristics of the equipment being used are recognized and appropriately compensated for.

### 2.2 Pole-Zero Modeling Approaches

A more complete model for speech is as an ARMA (autoregressive moving average) process with both poles and zeros:

$$s(n) = \sum_{i=1}^{L} b_i s_{n-i} + \sum_{j=1}^{M} a_j g_{n-j} + g(n) \ . \tag{9}$$

Such a model allows for more realistic modeling of speech sounds apart from vowels, particularly nasals, fricatives and stop consonants [58]. However, estimating the parameters of a pole-zero model is a nonlinear estimation problem [47]. There are many different approaches to the estimation of a pole-zero model for speech ranging from inverse LPC [47], iterative pre-filtering [72], [73], SEARMA (simultaneous estimation of ARMA parameters) [58], weighted recursive least squares (WRLS) [29], [54], [55], weighted least squares lattice [45], WRLS with variable forgetting factor (WRLS-VFF) [18]. These methods can give very good results but are computationally more intensive. They have the advantage that they can easily be extended to track the time-varying characteristics of speech [18],[54],[77], but the limited amount of data can lead to problems with convergence. Parametric techniques also have stability problems when the model order is not estimated correctly [58].

The periodic nature of voiced speech is a difficulty [55] which may be dealt with by incorporating simultaneous estimation of the input [54], [55]. If the input

is assumed to be either a pseudo-periodic pulse train or white noise, the pole-zero model obtained will include the lip radiation, the vocal tract filter and the glottal waveform and there is no obvious way to separate the poles and zeros which model these different features [54].

ARMA modeling approaches have been used to perform closed phase glottal pulse inverse filtering [77] giving advantages over frame-based techniques such as linear prediction by eliminating the influence of the pitch, leading to better accuracy of parameter estimation and better spectral matching [77]. In [46],[77], WRLS-VFF is used to perform closed phase glottal pulse inverse filtering and the variable forgetting factor is used to predict the presence or absence of the glottal closed phase which then allows for a more accurate estimate of the formants and anti-formants. The main drawbacks of the approach [77] are computational complexity and the difficulty of obtaining good a priori information on model order and model type i.e. the relative number of poles and zeros.

**Model Based Approaches.** As seen above, it is possible to develop time-varying pole-zero models of speech, but, if the input is modelled as a pulse train or white noise, it is not possible unambiguously to determine which poles and zeros model the glottal source excitation. Only by a combination of adaptive ARMA modelling and inverse filtering such as in [77] is it then possible to recover the glottal waveform. An extension of pole-zero modelling to include a model of the glottal source excitation can overcome the drawbacks of inverse filtering and produces a parametric model of the glottal waveform. In [43], the glottal source is modelled using the LF model [27] and the vocal tract is modelled as two distinct filters, one for the open phase, one for the closed phase [63]. Glottal closure is identified using the EGG. In [30,31] the LF model is also used in adaptively and jointly estimating the vocal tract filter and glottal source using Kalman filtering. To provide robust initial values for the joint estimation process, the problem is first solved in terms of the Rosenberg model [66]. One of the main drawbacks of model-based approaches is the number of parameters which need to be estimated for each period of the signal [43] especially when the amount of data is small e.g. for short pitch periods in higher voices. To deal with this problem, inverse filtering may be used to remove higher formants and the estimates can be improved by using ensemble averaging of successive pitch periods.

Modeling techniques need not involve the use of standard glottal source models. Fitting polynomials to the glottal wave shape is a more flexible approach which can place fewer constraints on the result. In [51], the differentiated glottal waveform is modelled using polynomials (a linear model) where the timing of the glottis opening and closing is the parameter which varies. Initial values for the glottal source endpoints plus the pitch period endpoints are found using the EGG. The vocal tract filter coefficients and the glottal source endpoints are then jointly estimated across the whole pitch period. This approach is an alternative to closed phase inverse filtering in the sense that even closed phase inverse filtering contains an implied model of the glottal pulse [51], i.e. the assumption of zero airflow through the glottis for the segment of speech from which the inverse filter coefficients are estimated. An alternative is to attempt to optimize

the inverse filter with respect to a glottal waveform model for the whole pitch period [51]. Interestingly in this approach, the result is the appearance of ripple in the source-corrected inverse filter during the closed phase of the glottal source even for synthesized speech with zero excitation during the glottal phase, (note that the speech was synthesized using the Ishizaka-Flanagan model [37]). Thus, this ripple must be an analysis artefact due to the inability of the model to account for it [51]. Improvements to the model are presented in [52],[76] and the sixth-order Milenkovic model is used in GELP [19].

In terms of the potential applications of glottal inverse filtering, the main difficulty with the use of glottal source models in glottal waveform estimation arises from the influence the models may have on the ultimate shape of the result. This is a particular problem with pathological voices. The glottal waveforms of these voices may diverge quite a lot from the idealized glottal models. As a result, trying to recover such a waveform using an idealized source model as a template may give less than ideal results. A model-based approach which partially avoids this problem is described in [64] where nonlinear least squares estimation is used to fit the LF model to a glottal derivative waveform extracted by closed phase filtering (where the closed phase is identified by the absence of formant modulation). This model-fitted glottal derivative waveform is the coarse structure. The fine structure of the waveform is then obtained by subtraction from the inverse filtered waveform. In this way, individual characteristics useful for speaker identification may be isolated. This approach also shows promise for isolating the characteristics of vocal pathologies.

### 2.3 Adaptive Inverse Filtering Approaches

For successful glottal waveform inverse filtering, an accurate vocal tract filter must first be acquired. In closed phase inverse filtering, the vocal tract filter impulse response is obtained free of the influence of the glottal waveform input. The influence of the glottal waveform can also be removed in the frequency domain. In the iterative adaptive inverse filtering method (IAIF-method) [5], a 2 pole model of the glottal waveform based on the characteristic 12dB/octave tilt in the spectral envelope [26] is used to remove the influence of the glottal waveform from the speech signal. The resulting vocal tract filter estimate is applied to the original speech signal to obtain a better estimate of the glottal waveform. The procedure is then repeated using a higher order parametric model of the glottal waveform. As the method removes the influence of the glottal waveform from the speech before estimating the vocal tract filter, it does not take a closed phase approach but utilises the whole pitch period. A flow diagram of the IAIF-method is shown in Fig. 4.

The method relies on linear prediction and is vulnerable to the deficiencies of that technique such as incorrect formant estimation due to the underlying harmonic structure in speech [47]. In particular, the technique performs less well for higher fundamental frequency voices [6]. To remove the influence of the pitch

period, the iterative adaptive procedure may be applied pitch synchronously [7] as shown in Fig. 5.

Comparing the results of the IAIF method with closed phase inverse filtering show that the IAIF approach seems to produce waveforms which have a shorter and rounder closed phase. In [7] comparisons are made between original and estimated waveforms for synthetic speech sounds. It is interesting to note that pitch synchronous IAIF produces a closed phase ripple in these experiments (when there was none in the original synthetic source waveform).

In [8] discrete all-pole modelling was used to avoid the bias given toward harmonic frequencies in the model representation. An alternative iterative approach is presented in [2]. The method de-emphasises the low frequency glottal information using high-pass filtering prior to analysis. In addition to minimising the influence of the glottal source, an expanded analysis region is provided in the form of a pseudo-closed phase. The technique then derives an optimum vocal tract filter function through applying the properties of minimum phase systems.

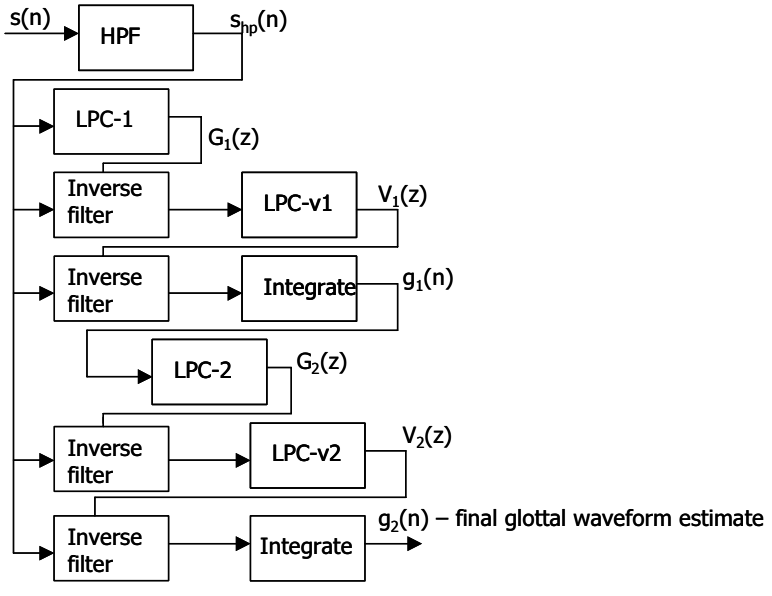



**Fig. 4.** The iterative adaptive inverse filtering method

**Other Iterative Approaches.** In [48] another iterative approach to glottal waveform estimation is developed. It is based on iterative inverse filtering (ITIF) [41], a technique for simultaneously estimating the poles and zeroes of an ARMA model based on the assumption that the input has a flat spectrum. In [48], the ITIF is used to find the poles and zeroes of a filter which will generate the

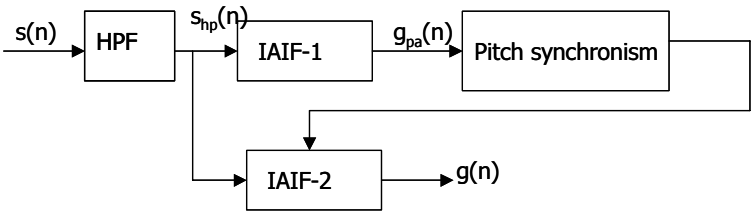



**Fig. 5.** The pitch synchronous iterative adaptive inverse filtering method

glottal waveform given an impulse train input. Doubly differentiating the speech production model gives:

$$(1 - z^{-1})S(z) = AV(z)(1 - z^{-1})Q(z) \tag{10}$$

where the doubly differentiated glottal waveform $Q(z)(1 - z^{-1})$, is just a periodic impulse train and so the estimation error may be approximated as the linear prediction residual. The inverse filter $I\,(z) = \frac{1}{V(z)}$ is determined using the covariance method and used to inverse filter the pre-emphasized speech. The residual is then integrated twice to yield the signal which has a spectrum which is an approximation of the glottal waveform amplitude spectrum [48] and from which a pole-zero glottal filter model may be determined using ITIF. The glottal filter may be used to generate an estimate of the glottal waveform when an input of a periodic impulse train is applied in the reverse time direction. The glottal waveforms so obtained typically have a negative-going closed phase, even when synthetic glottal waveforms with closed phase equal to zero are recovered. Typically, the models used in the glottal filter in this work have 14 zeros and 2 poles. However, it has been suggested [49], that the glottal waveform can be modelled purely by equally spaced zeros. Interestingly, in [48], an improved result is found when an all-zero filter is developed, where as many as 26 zeros may be required.

### 2.4 Higher Order Statistics and Cepstral Approaches

These approaches exploit the additional properties of new statistical techniques. For example, higher order statistics such as the bispectrum (third-order spectrum) are theoretically immune to Gaussian noise (in practice there is always some noise because of fixed length data records) [50], [59]. The bispectrum also contains system phase information and many bispectrum-based blind deconvolution algorithms have been developed to recover any type of system including non-minimum phase systems for a non-Gaussian white input. By assuming the pseudo-periodic pulse train as input (non-Gaussian white noise) the periodic aspect of the speech is assumed to be accounted for, but this is not necessarily the case. The main drawback with bispectral and other higher order statistics approaches is that they require greater amounts of data to reduce the variance in the spectral estimates [35]. As a result, multiple pitch periods are required which would necessarily be pitch asynchronous. This problem may be overcome

by using the Fourier series and thus performing a pitch synchronous analysis [34]. It has been demonstrated that the higher order statistics approach can recover a system filter for speech, particularly for speech sounds such as nasals [34]. Such a filter may be non-minimum phase and when its inverse is used to filter the speech signal will return a residual which is much closer to a pure pseudo-periodic pulse train than inverse filters produced by other methods [14], [34]. In [14], the speech input estimate generated by this approach is used in a second step of ARMA parameter estimation by an input-output system identification method.

The properties of the cepstrum have also been exploited in speech processing. Transformed into the cepstral domain, the convolution of input pulse train and vocal tract filter becomes an addition of disjoint elements, allowing the separation of the filter from the harmonic component [61]. Cepstral techniques also have some limitations including the requirement for phase unwrapping and the fact that the technique cannot be used (although it often is) when there are zeros on the unit circle. In [42], various ARMA parameter estimation approaches are applied to the vocal tract impulse response recovered from the cepstral analysis of the speech signal [60].

There are a few examples of direct glottal waveform recovery using higher order spectral or cepstral techniques. In [80], ARMA modelling of the linear bispectrum [25] was applied to speech for joint estimation of the vocal tract model and the glottal volume velocity waveform using higher-order spectral factorization [75] with limited success. Direct estimation from the complex cepstrum was used in [4] based on the assumption that the glottal volume velocity waveform may be modelled as a maximum phase system. As the complex cepstrum separates into causal and acausal parts corresponding to the minimum and maximum phase parts of the system model this then permits a straightforward separation of the glottal waveform.

## 3 Effect of Recording Conditions

With a fundamental frequency varying in the range 80–250 Hz the glottal waveform is a low-frequency signal and so the low-frequency response, including the phase response, of the recording equipment used is an important factor in glottal pulse identification. However, many authors do not report in detail on this. In [81], the following potential problems are identified: ambient noise, low-frequency bias due to breath burst on the microphone, equipment and tape distortion of the signal and improper A/D conversion (p. 355). The problem of ambient noise can be overcome by ensuring suitable recording conditions. Use of special equipment [67], [70] can also minimize the noise problem, but is not always possible, or may not be relevant to the method under investigation. Paradoxically, the problem of the low-frequency bias producing a trend in the final recovered waveform can occur when a high quality microphone and amplifier with a flat frequency response down to 20 Hz are used. It can be overcome by high pass filtering with a cut-off frequency no greater than half the fundamental frequency [81] or by cancelling out the microphone transfer function [51], [79].

Phase distortion was a problem with analog tape recording [12], [36] and is illustrated as a characteristic 'humped' appearance, although doubtless there are other causes as such waveforms are still being recovered [43]. For example, phase distortion can result from HOS and cepstral approaches when phase errors occur due to the need for phase unwrapping [80]. However, more modern recording techniques especially involving the use of personal computer (PC) sound cards can also introduce phase distortion at low frequencies which will impact on the glottal waveform reconstruction. This is clearly demonstrated by experiments conducted by [1] where synthetic glottal waveforms created using the LF model were recorded through a PC sound card (Audigy2 SoundBlaster) resulting in the characteristic humped appearance. The effect was noticeable up to 320 Hz, but was especially pronounced at the lowest fundamental frequency (80 Hz). In all cases, the flat closed phase was entirely lost. The correction technique proposed is to model the frequency response of the recording system using a test signal made of a sum of sinusoids and thus to develop a compensating filter [1].

Few researchers take the care shown in [79] who plots an example of a glottal waveform with a widely varying baseline due to the influence of low-frequency noise picked up by a high-quality microphone such as a Brüel & Kjær 4134 [6], [79]. To overcome this problem, it is common to high-pass filter the speech [6], [42], [81] but according to [79] this is not sufficient as it removes the flat part of the closed phase and causes an undershoot at glottal closure: a better approach is to compensate by following the high pass filter by a low pass filter. According to [81], a pole may arise at zero frequency due to a non-zero mean in the typically short duration closed phase analysis window. It appears that in [81] such a pole is removed from the inverse filter if it arises (and not by linear phase high pass filtering as suggested by [79]), whereas in [79] the resulting bias is removed by polynomial fitting to 'specific points of known closed phase' (presumably the flattest points). An alternative approach is to take advantage of specialized equipment such as the Rothenberg mask [67] to allow for greater detail of measurement of the speech signal at low frequencies. The characteristics of the mask may then be removed by filtering during analysis [51].

Most experimenters who have reported on recording conditions have used condenser type microphones [6], [79], [81] with the exception of [51] who claims that these microphones are prone to phase distortion around the formant frequencies. However, available documentary information on microphone characteristics [13], the weight of successful inverse filtering results using condenser microphone recordings and direct comparison of results with different microphone types [15], [64] seem to contradict this claim. Depending on the application, it will not always be possible to apply such stringent recording conditions. For example, Plumpe et al. [64] test a glottal flow based speaker identification on samples from the TIMIT and NTIMIT databases. The TIMIT database is recorded with a high-quality (Sennheiser) microphone in a quiet room while the NTIMIT database represents speech of telephone-channel quality. Here it is in fact the cheaper microphone which is suspected of causing phase distortion which shows up in the estimated glottal flow derivatives. In other cases, the recording conditions

may not be under the control of the investigator who may be using commercially provided data sources such as [16].

## 4 Evaluation of Results

One of the primary difficulties in glottal pulse identification is in the evaluation of the resulting glottal flow waveforms. How do we know we have the 'right answer'? How do we even know what the 'right answer' looks like? There are several approaches which can be taken. One approach is to verify the algorithm which is being used for the glottal flow waveform recovery. Algorithms can be verified by applying the algorithm to a simulated system which may be synthesized speech but need not be [41], [42]. In the case of synthesized speech, the system will be a known all-pole vocal tract model and the input will be a model for a glottal flow waveform. The success of the algorithm can be judged by quantifying the error between the known input waveform and the version recovered by the algorithm. This approach is most often used as a first step in evaluating an algorithm [6], [7], [48], [77], [80] and can only reveal the success of the algorithm in inverse filtering a purely linear time-invariant system. Synthesized speech can also be provided to the algorithm using a more sophisticated articulatory model [37] which allows for source-tract interaction [51].

Once an algorithm has been verified and is being used for inverse filtering real speech samples, there are two possible approaches to evaluating the results. One is to compare the waveforms obtained with those obtained by other (usually earlier) approaches. As, typically, the aim of this is to establish that the new approach is superior, the objectivity of this approach is doubtful. This approach can be made most objective when methods are compared using synthetic speech and results can be compared with the original source, as in [7]. However, the objectivity of this approach may also be suspect because the criteria used in the comparison are often both subjective and qualitative as for example in [77] where visual inspection seems to be the main criterion: "The WRLS-VFF method appears to agree with the expected characteristics for the glottal excitation source such as a flat closed region and a sharp slope at closure better than the other two methods." (p. 392) Other examples of such comparisons are in [24] and [43]. In many papers no comparisons are made, a stance which is not wholly unjustified because there is not a great deal of data available to say which are the correct glottal flow waveforms.

On the other hand, using two different methods to extract the glottal flow could be an effective way to confirm the appearance of the waveform as correct. The rationale behind this is that if two (or more) different approaches garner the same result then it has a greater chance of being 'really there'. If one of the methods, at least for experimental work, utilizes additional help such as the EGG to accurately identify glottal closure, then that would provide additional confirmation. This approach was taken in [43] but the results, albeit similar for two approaches, are most reminiscent of a type of waveform labelled as exhibiting phase distortion in

[81]. The same could be said about many of the results offered in [24] and [80], where low-frequency (baseline) drift is also in evidence. Once again, if new techniques for glottal inverse filtering produce waveforms which 'look like' the other waveforms which have been produced before, then they are evaluated as better than those which do not: examples of the latter include [4], [22].

Improved guidelines for assessing glottal waveform estimates can come from experiments with physiologically based articulatory synthesis methods. Glottal inverse filtering can be applied to speech produced with such models where the models are manipulated to produce various effects. The types of glottal waveforms recovered can then be assessed in the light of the perturbations introduced. An interesting example of what is possible with this idea is shown by [20] where various degrees and types of air leakage are shown to correlate with varying amounts of open and closed phase ripple in the derivative glottal flow and the glottal flow itself.

An alternative approach is to apply some objective mathematical criterion. In [23], it is shown how the evolution of the phase-plane plot of $g(t)$ versus $\frac{dg(t)}{dt}$ to a single closed loop indicates that a periodic solution has been produced and all resonances have been removed since resonances will appear as self-intersecting loops on the phase-plane plot.

### 4.1 Separability of Tract and Source

Glottal source models based on the linearly separable speech production model [27], [39], [40], [66], and derived from these early studies are still very successfully used in speech coding and speech synthesis [21]. Most of these models, while not as simple as the periodic impulse train, are relatively simple to generate, while the more complex models such as the LF model [27] produce the best results and have the added advantage of being a model of the derivative glottal flow and so automatically include lip radiation [21]. The method cannot be used where the actual speech production does not fit the model, for example in higher pitched voices (females, children) where the glottis does not close completely.

According to [44], the vocal tract filter is separable from the source only if the source itself is correctly defined. It has been shown that source-tract interaction can affect the glottal waveform [44] including the appearance of a first formant ripple on the waveform. There are effectively two ways of achieving this separation [9]: either assume the source is independent and have a time-varying vocal tract filter which will have different formants and bandwidths in closed and open phases or define the source as derived from the closed phase vocal tract as the true source and assume the vocal tract filter is time-invariant. Using the second solution, the variation in the formant frequency and bandwidth has to go somewhere and it ends up as a ripple on the open phase part of the glottal volume velocity (see for example Fig. 5c in [81]). Thus, strictly speaking, due to source-tract interaction, linear prediction analysis applied to a whole pitch period will contain slight formant frequency and bandwidth errors [44]. Also, according to this definition, a 'true' glottal volume velocity waveform can only

be obtained by inverse filtering by a closed phase method and it should have the ripple (more visible on the differentiated waveform) and a flat closed phase.

However, a common result in inverse filtering is a ripple in the closed phase of the glottal volume velocity waveform. In [79] this occurs in hoarse or breathy speech and is assumed to show that there is air flow during the glottal closed phase. In [79] it is shown through experiments that this small amount of air flow does not significantly alter the inverse filter coefficients (filter pole positions change by $< 4\%$) and that true non-zero air flow can be captured in this way. However, the non-zero air flow and resultant source-tract interaction may still mean that the 'true' glottal volume velocity waveform is not exactly realized [79]. A similar effect is observed when attempting to recover source waveforms from nasal sounds. Here the strong vocal tract zeros mean that the inverse filter is inaccurate and so a strong formant ripple appears in the closed phase [79].

Most recently, a sliding window approach to closed phase inverse filtering has been attempted [21], [24]. Originally this approach required manual intervention to choose the best glottal waveform estimates from those obtained in periods of glottal closure in the speech waveform which were also identified by the operator [21]. Again, this is a very subjective procedure. Such an approach may be automated by using the maxiumum amplitude negative peaks in the linear prediction residual to estimate the glottal closure, but this is nothing new [81]. The best glottal waveform estimates are also chosen automatically by choosing the smoothest estimates [24]. The results obtained by this method were verified by comparing with waveforms obtained using the EGG to detect glottal closure.

## 5 Conclusion

Although convincing results for glottal waveform characteristics are reported in the literature from time to time, a standard fully automatic inverse filtering algorithm is not yet available. An extensive review of the literature has established that the salient features of the glottal waveform were established fairly early on, as was the technique of choice which continues to be closed phase inverse filtering. This technique has been successful because it allows the adoption of the linear time-invariant model for both determining the filter in the source-filter speech model and then for applying it as an inverse filter to recover the source. Despite concern about features of recovered waveforms which may be due to inaccuracies and oversimplifications in this model, alternative approaches have met with limited success. ARMA modelling has limitations due to the insufficiency of data and the 'magic bullet' promise of alternative statistical techniques such as the cepstrum and higher order statistics has not delivered. Low frequency phase response and low frequency noise have been shown to be important issues for glottal waveform recovery (at least in some contexts such as vocal pathology, experimental studies on voice production and benchmark generation) which have not always received due attention by researchers. However, nonlinear approaches (with the exception of the statistical techniques mentioned already) are only just beginning to be explored.

## References


1. Akande, O., O.: Speech analysis techniques for glottal source and noise estimation in voice signals. Ph. D. Thesis, University of Limerick (2004)
2. Akande, O. and Murphy, P. J.: Estimation of the vocal tract transfer function for voiced speech with application to glottal wave analysis. Speech Communication, **46** (2005) 15–36
3. Akande, O., Murphy, P. J.: Improved speech analysis for glottal excited linear predictive speech coding. Proc. Irish Signals and Systems Conference. (2004) 101–106
4. Alkhairy, A.: An algorithm for glottal volume velocity estimation. Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing. **1** (1999) 233–236
5. Alku, P., Vilkman, E., Laine, U. K.,: Analysis of glottal waveform in different phonation types using the new IAIF-method. Proc. 12th Int. Congress Phonetic Sciences, **4** (1991) 362–365
6. Alku, P.: An automatic method to estimate the time-based parameters of the glottal pulseform. Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing. **2** (1992) 29–32
7. Alku, P.: Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. Speech Communication. **11** (1992) 109–118
8. Alku, P., Vilkman, E.: Estimation of the glottal pulseform based on Discrete All-Pole modeling. Proc. Int. Conf. on Spoken Language Processing. (1994) 1619-1622
9. Ananthapadmanabha, T. V., Fant, G.: Calculation of true glottal flow and its components. STL-QPR. (1985) 1–30
10. Atal, B. S., Hanauer, S. L.: Speech analysis and synthesis by linear prediction of the speech wave. J. Acoust. Soc. Amer. **50** (1971) 637–655
11. Bergstrom, A., Hedelin, P.: Codebook driven glottal pulse analysis. Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing. **1** (1989) 53–56
12. Berouti, M., Childers, D., Paige, A.: Correction of tape recorder distortion. Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing. **2** (1977) 397–400
13. Brüel & Kjær: Measurement Microphones. 2nd ed. (1994)
14. Chen, W.-T., Chi, C.-Y.: Deconvolution and vocal-tract parameter estimation of speech signals by higher-order statistics based inverse filters. Proc. IEEE Workshop on HOS. (1993) 51–55
15. Childers, D. G.: Glottal source modeling for voice conversion. Speech Communication. **16** (1995) 127–138
16. Childers, D. G.: Speech processing and synthesis toolboxes. Wiley: New York (2000)
17. Childers, D. G., Chieteuk, A.: Modeling the glottal volume-velocity waveform for three voice types. J. Acoust. Soc. Amer. **97** (1995) 505–519
18. Childers, D. G., Principe, J. C., Ting, Y. T. Adaptive WRLS-VFF for Speech Analysis. IEEE Trans. Speech and Audio Proc. **3** (1995) 209–213
19. Childers, D. G., Hu, H. T.: Speech synthesis by glottal excited linear prediction. J. Acoust. Soc. Amer. **96** (1994) 2026-2036
20. Cranen, B., Schroeter, J.: Physiologically motivated modelling of the voice source in articulatory analysis/synthesis. Speech Communication. **19** (1996) 1–19
21. Cummings, K. E., Clements, M. A.: Glottal Models for Digital Speech Processing: A Historical Survey and New Results. Digital Signal Processing. **5** (1995) 21–42
22. Deng, H., Beddoes, M. P., Ward, R. K., Hodgson, M.: Estimating the Glottal Waveform and the Vocal-Tract Filter from a Vowel Sound Signal. Proc. IEEE Pacific Rim Conf. Communications, Computers and Signal Processing. **1** (2003) 297–300

23. Edwards, J. A., Angus, J. A. S.: Using phase-plane plots to assess glottal inverse filtering. Electronics Letters **32** (1996) 192–193
24. Elliot, M., Clements, M.: Algorithm for automatic glottal waveform estimation without the reliance on precise glottal closure information. Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing. **1** (2004) 101–104
25. Erdem, A. T., Tekalp, A. M.: Linear Bispectrum of Signals and Identification of Nonminimum Phase FIR Systems Driven by Colored Input. IEEE Trans. Signal Processing. **40** (1992) 1469–1479
26. Fant, G. C. M.: Acoustic Theory of Speech Production. (1970) The Hague, The Netherlands: Mouton
27. Fant, G., Liljencrants, J., Lin, Q.: A four-parameter model of glottal flow. STL-QPR. (1985) 1–14
28. Fant, G., Lin, Q., Gobl, C.: Notes on glottal flow interaction. STL-QPR. (1985) 21–45
29. A recursive maximum likelihood algorithm for ARMA spectral estimation. IEEE Trans. Inform. Theory **28** (1982) 639–646
30. Fu, Q., Murphy, P.: Adapive Inverse filtering for High Accuracty Estimation of the Glottal Source. Proc. NoLisp'03. (2003)
31. Fu, Q., Murphy, P. J.: Robust glottal source estimation based on joint source-filter model optimization. IEEE Trans. Audio, Speech Lang. Proc., **14** (2006) 492–501
32. Hillman, R. E., Weinberg, B.: A new procedure for venting a reflectionless tube. J. Acoust. Soc. Amer. **69** (1981) 1449–1451
33. Holmberg, E. R., Hillman, R. E., Perkell, J. S.: Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice. J. Acoust. Soc. Amer. **84** (1988) 511–529
34. Hinich, M. J., Shichor, E.: Bispectral Analysis of Speech. Proc. 17th Convention of Electrical and Electronic Engineers in Israel. (1991) 357–360
35. Hinich, M. J., Wolinsky, M. A.: A test for aliasing using bispectral components. J. Am. Stat. Assoc. **83** (1988) 499-502
36. Holmes, J. N.: Low-frequency phase distortion of speech recordings. J. Acoust. Soc. Amer. **58** (1975) 747–749
37. Ishizaka, K., Flanagan, J. L.: Synthesis of voiced sounds from a two mass model of the vocal cords. Bell Syst. Tech. J. **51** (1972) 1233–1268
38. Jiang, Y., Murphy, P. J.: Production based pitch modification of voiced speech. Proc. ICSLP, (2002) 2073–2076
39. Klatt, D.: Software for a cascade/parallel formant synthesizer. J. Acoust. Soc. Amer. **67** (1980) 971–994
40. Klatt, D., Klatt, L.: Analysis, synthesis, and perception of voice quality variations among female and male talkers. J. Acoust. Soc. Amer. **87** (1990) 820–857
41. Konvalinka, I. S., Mataušek, M. R.: Simultaneous estimation of poles and zeros in speech analysis and ITIT-iterative inverse filtering algorithm. IEEE Trans. Acoust., Speech, Signal Proc. **27** (1979) 485–492
42. Kopec, G. E., Oppenheim, A. V., Tribolet, J. M.: Speech Analysis by Homomorphic Prediction IEEE Trans.Acoust., Speech, Signal Proc. **25** (1977) 40–49
43. Krishnamurthy, A. K.: Glottal Source Estimation using a Sum-of-Exponentials Model. IEEE Trans. Signal Processing. **40** (1992) 682–686
44. Krishnamurthy, A. K., Childers, D. G.: Two-channel speech analysis. IEEE Trans. Acoust., Speech, Signal Proc. **34** (1986) 730–743
45. Lee, D. T. L., Morf, M., Friedlander, B.: Recursive least squares ladder estimation algorithms. IEEE Trans. Acoust., Speech, Signal Processing. **29** (1981) 627–641

46. Lee, K., Park, K.: Glottal Inverse Filtering (GIF) using Closed Phase WRLS-VFF-VT Algorithm. Proc. IEEE Region 10 Conference. **1** (1999) 646–649
47. Makhoul, J.: Linear Prediction: A Tutorial Review. Proc. IEEE. **63** (1975) 561–580
48. Mataušek, M. R., Batalov, V. S.: A new approach to the determination of the glottal waveform. IEEE Trans. Acoust., Speech, Signal Proc. **28** (1980) 616–622
49. Mathews, M. V., Miller, J. E., David, Jr., E. E.: Pitch synchronous analysis of voiced sounds. J. Acoust. Soc. Amer. **33** (1961) 179–186
50. Mendel, J. M.: Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications. Proc. IEEE. **79** (1991) 278–305
51. Milenkovic, P.: Glottal Inverse Filtering by Joint Estimation of an AR System with a Linear Input Model. IEEE Trans. Acoust., Speech, Signal Proc. **34** (1986) 28–42
52. Milenkovic, P. H.: Voice source model for continuous control of pitch period. J. Acoust. Soc. Amer. **93** (1993) 1087-1096
53. Miller, R. L.: Nature of the Vocal Cord Wave. J. Acoust. Soc. Amer. **31** (1959) 667-677
54. Miyanaga, Y., Miki, M., Nagai, N.: Adaptive Identification of a Time-Varying ARMA Speech Model. IEEE Trans. Acoust., Speech, Signal Proc. **34** (1986) 423–433
55. Miyanaga, Y., Miki, N., Nagai, N., Hatori, K.: A Speech Analysis Algorithm which eliminates the Influence of Pitch using the Model Reference Adaptive System. IEEE Trans. Acoust., Speech, Signal Proc. **30** (1982) 88–96
56. Monsen, R. B., Engebretson, A. M.: Study of variations in the male and female glottal wave. J. Acoust. Soc. Amer. **62** (1977) 981–993
57. Monsen, R. B., Engebretson, A. M., Vemula, N. R.: Indirect assessment of the contribution of subglottal air pressure and vocal-fold tension to changes of fundamental frequency in English. J. Acoust. Soc. Amer. **64** (1978) 65–80
58. Morikawa, H., Fujisaki, H.: Adaptive Analysis of Speech based on a Pole-Zero Representation. IEEE Trans. Acoust., Speech, Signal Proc. **30** (1982) 77–87
59. Nikias, C. L., Raghuveer, M. R.: Bispectrum Estimation:A Digital Signal Processing Framework. Proc. IEEE. **75** (1987) 869–891
60. Oppenheim, A. V.: A speech analysis-synthesis system based on homomorphic filtering. J. Acoust., Soc. Amer. **45** (1969) 458–465
61. Oppenheim, A. V., Schafer, R. W.: Discrete-Time Signal Processing. Englewood Cliffs:London Prentice-Hall (1989)
62. Pan, R., Nikias, C. L.: The complex cepstrum of higher order cumulants and nonminimum phase system identification. IEEE Trans. Acoust., Speech, Signal Proc. **36** (1988) 186–205
63. Parthasarathy, S., Tufts, D. W.: Excitation-Synchronous Modeling of Voiced Speech. IEEE Trans. Acoust., Speech, Signal Proc. **35** (1987) 1241–1249
64. Plumpe, M. D., Quatieri, T. F., Reynolds, D. A.: Modeling of the Glottal Flow Derivative Waveform with Application to Speaker Identification. IEEE Trans. Speech and Audio Proc. **7** (1999) 569–586
65. Quatieri, T. F., McAulay, R. J.: Shape invariant time-scale and pitch modification of speech. IEEE Trans. Signal Process., **40** (1992) 497–510
66. Rosenberg, A.: Effect of the glottal pulse shape on the quality of natural vowels. J. Acoust. Soc. Amer. **49** (1971) 583–590
67. Rothenberg, M.: A new inverse-filtering technique for deriving the glottal air flow waveform. J. Acoust. Soc. Amer. **53** (1973) 1632–1645

68. Schroeder, M. R., Atal, B. S.: Code-excited linear prediction (CELP): High quality speech at very low bit rates. Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing. **10** (1985) 937–940
69. Shanks, J. L.: Recursion filters for digital processing. Geophysics. **32** (1967) 33–51
70. Sondhi, M. M.: Measurement of the glottal waveform. J. Acoust. Soc. Amer. **57** (1975) 228–232
71. Sondhi, M. M., Resnik, J. R.: The inverse problem for the vocal tract: Numerical methods, acoustical experiments, and speech synthesis. J. Acoust. Soc. Amer. **73** (1983) 985–1002
72. Steiglitz, K.: On the simultaneous estimation of poles and zeros in speech analysis. IEEE Trans. Acoust., Speech, Signal Proc. **25** (1977) 194–202
73. Steiglitz, K., McBride, L. E.: A technique for the identifcation of linear systems. IEEE Trans. Automat. Contr., **10** (1965) 461–464
74. Stylianou, Y.: Applying the harmonic plus noise model in concatenative speech synthesis. IEEE Trans. Speech Audio Process., **9**(2001) 21–29
75. Tekalp, A. M., Erdem, A. T.: Higher-Order Spectrum Factorization in One and Two Dimensions with Applications in Signal Modeling and Nonminimum Phase System Identification. IEEE Trans. Acoust., Speech, Signal Proc. **37** (1989) 1537–1549
76. Thomson, M. M.: A new method for determining the vocal tract transfer function and its excitation from voiced speech. Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing. **2** (1992) 23–26
77. Ting, Y., T., Childers, D. G.: Speech Analysis using the Weighted Recursive Least Squares Algorithm with a Variable Forgetting Factor. Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing. **1** (1990) 389–392
78. Tremain, T. E.: The government standard linear predictive coding algorithm: LPC-10. Speech Technol. 1982 40–49
79. Veeneman, D. E., BeMent, S. L.: Automatic Glottal Inverse Filtering from Speech and Electroglottographic Signals. IEEE Trans. Acoust., Speech, Signal Proc. **33** (1985) 369–377
80. Walker, J.: Application of the bispectrum to glottal pulse analysis. Proc. NoLisp'03. (2003)
81. Wong, D. Y., Markel, J. D., Gray, A. H.: Least squares glottal inverse filtering from the acoustic speech waveform. IEEE Trans. Acoust., Speech, Signal Proc. **27** (1979) 350–355