ISCA Archive
http://www.isca-speech.org/archive

ITRW on Non-Linear Speech Processing (NOLISP 03)
Le Croisic, France
May 20-23, 2003

Application of the Bispectrum to Glottal Pulse Analysis

Dr Jacqueline Walker
Department of Electronic and Computer Engineering
University of Limerick
jacqueline.walker@ul.ie

***Abstract*-Higher order spectral (HOS) techniques, such as the bispectrum, offer robustness to Gaussian noise and the ability to recover phase information. However, their drawbacks, such as the high variance of estimates and the need for long data records, have limited their use in conventional speech processing applications. As in glottal pulse estimation, all existing inverse filtering approaches use second-order statistics, it is of interest to explore the potential of HOS in this area. Using the theory of HOS factorization and the linear bispectrum, it is shown how voiced speech can be modelled as a nonGaussian coloured noise driven system. The linear bispectrum approach can be used to obtain alternative glottal pulse and vocal tract estimates in hybrid Iterative Adaptive Inverse Filtering (hIAIF) and the results are compared with traditional IAIF. Finally, a new technique which involves joint estimation of the glottal pulse and vocal tract followed by inverse filtering is presented. This new technique shows good preliminary results and is much simpler than previous techniques.**

I INTRODUCTION

Higher order spectral (HOS) techniques have been the subject of considerable research interest for the last 15 years. Although not quite living up to their original billing [1], [2], there has been development in the field, particularly in the areas of robust signal reconstruction [3], [4] and system identification [5], [6]. The attraction of higher order spectral techniques rests in their apparent noise-suppression properties and their retention of system phase information.

The main drawback with the use of HOS is that, to obtain reliable estimates, large data records are needed [7]. Frequency domain smoothing [7] or time domain segment averaging [8] are used to reduce the variance of the estimates. As speech is generally regarded as approximately stationary in frames of 10-30ms, obtaining a sufficiently long data record which can provide adequate frequency resolution is difficult [9]. Conversely, attempts to apply HOS to short data records often provide discouraging results [10] and this could partly explain why HOS approaches have not been pursued significantly in speech signal processing.

Below some examples of the application of HOS techniques, notably the bispectrum, in speech processing applications are presented, followed by an overview of currently used techniques for glottal pulse estimation. In Section II, the theory of higher order spectral factorization [11] and the linear bispectrum [12] are presented and it is shown how the linear bispectrum might be used to extract the glottal pulse from the speech signal. In Section III linear bispectrum based estimation of system parameters is applied in a new algorithm called hybrid Iterative Adaptive Inverse Filtering (hIAIF). Results from this approach are presented for comparison with results from traditional Iterative Adaptive Inverse Filtering (IAIF) [13]. Finally, a new technique which involves joint estimation of the glottal pulse and vocal tract followed by inverse filtering is presented. This new technique shows good preliminary results and is much simpler than previous iterative techniques as it is a two stage process.

A. Application of the Bispectrum in Characterisation of Speech Production

To characterise speech production, a model needs to be assumed and two are presented in this section. It is shown how the two models relate to the bispectral approaches presented. The relationship between the two models and different types of speech coders is also briefly discussed. The performance of the coders serves to highlight where the models do or do not correspond with the reality of speech production. Finally, a brief overview of the various techniques which have been proposed specifically for glottal pulse estimation is given.

The simplest speech production model is shown in Fig. 1. Voiced speech is produced when impulses with the required pitch period are input to a linear system which models the vocal tract. Unvoiced speech is produced when the input to the system is white noise. The vocal tract model is usually estimated as a minimum-phase all-pole model using linear prediction. Such a model is used by the classical LPC (linear prediction coding) vocoder [14], which can code speech at very low bit rates. The reconstructed speech often sounds unnatural and has a ‘buzzy’ character which makes it hard to identify the speaker [15] suggesting that information has been lost. Hybrid speech coders such as CELP [16] overcome this problem by using an excitation codebook which effectively codes the residual signal by selecting the excitation which minimises the error between the reconstructed and original speech.

Application of HOS techniques to extract the parameters of this speech production model has two main motivations. Firstly, as the higher order spectra of Gaussian processes are theoretically zero, HOS techniques for vocal tract model parameter estimation are potentially more robust in the presence of additive Gaussian noise. Secondly, as HOS retain phase information, more information may be captured in vocal tract model parameters estimated using HOS techniques.

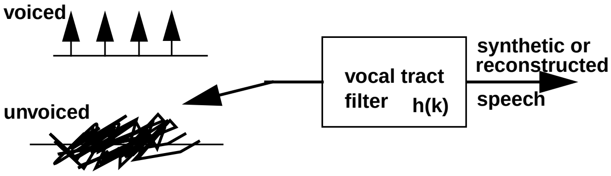

Fig. 1: Simple voiced/unvoiced speech model.

Both these motivations were present in the work presented in [10], which considered the use of a third-order statistics (TOS) based approach to AR parameter estimation which could be applied in speech coding. The AR parameters of the all-pole vocal tract model, which could be non-minimum phase, were calculated using a third-order cumulant-based linear prediction analysis [17]. It was found that the TOS approach gave better estimates than traditional linear prediction for noisy synthetic speech; for noiseless synthetic speech the results were about the same [10]. However, when used to code real speech using VSELP [18] the performance was no better, and in the case of noiseless speech, the results were worse. This degradation of performance was thought to be due to the higher variance in the TOS-based estimates due to the short-term analysis required.

In [19] the aim was to show that a better estimate of the speech signal waveform could be made using the bispectrum than with traditional techniques. A non-parametric bispectrum-based estimation was applied to voiced speech (a vowel 'A' and a nasal 'N'), where the bispectrum was estimated pitch synchronously using the Fourier series coefficients. The amplitude ($A(u)$) and phase ($\phi(u)$) of the vocal tract system filter were found from the bispectrum using established techniques [20] and the inverse system filter was then found as

$$H(u) = \frac{1}{A(u)}\exp(-j\phi(u)) \tag{1}$$

Applying this filter to speech produced a residual much closer to a sequence of pseudoperiodic impulses than with linear prediction techniques. From this, the authors concluded that the vocal tract model had been more accurately estimated. The analysis was also successfully performed for real speech in the presence of noise.

A second, more detailed model of speech production includes the glottal pulse, as shown in Fig. 2. The repetitive glottal pulses produced by the convolution of the pitch-period impulse train and the glottal pulse model $g(k)$ are input to the vocal tract model to generate voiced speech. Unvoiced speech is produced when the input to the vocal tract model is white noise. Obtaining explicit information about the glottal pulse has mainly been of interest in biomedical applications, such as vocal cord pathology investigations. More recently, however, new speech synthesis and coding techniques such as Glottal Excited Linear Prediction (GELP) [21], [22] have been proposed. Higher quality, more natural sounding speech should be obtainable by recovering and encoding features of the glottal excitation [22].

In [23], a prediction error system identification method was used to find the ARMA parameters of the non-minimum phase vocal tract system filter $h(k)$, once an estimate of its inverse $\hat{f}(k)$ was found by established third-order spectral techniques [24]. Once again, inverse filtering showed that the residual was very close to the ideal pseudoperiodic impulse train. Here the authors point out that their estimate of $h(k)$ must also include the 'vocal cord sound pulse' or glottal pulse but do not provide an approach to extracting this from the ARMA model of $h(k)$.

B. Glottal Pulse Estimation by Inverse Filtering

The main approach to determining the glottal pulse is inverse filtering, in which an estimate of the vocal tract filter is first made. By filtering the speech with the inverse of the vocal tract filter, the glottal excitation may be recovered.

In some inverse filtering techniques, the results are fitted to an established glottal pulse model such as the LF model [25] but this approach is limited in its applicability, especially to voice pathologies which would result in significant variations from the standard represented by such a model. In the closed phase approach [26], the vocal tract estimate is made only during the closed phase, when the glottal excitation may be assumed to be zero. The main problem with this approach is that vocal tract characteristics change between the closed and open phases - a problem which is not however, directly addressed by any of the competing techniques.

The most successful technique to date is IAIF [13] and its successor pitch-synchronous iterative adaptive inverse filtering (PS-IAIF) [27], where the vocal tract estimate is improved over iterations by first estimating and removing the effects of the glottal source. Inverse filtering using the refined vocal tract model produces the glottal pulse estimates.

All the inverse filtering approaches described above use second-order statistics based techniques which cannot preserve phase information. It therefore remains of interest to explore the potential of HOS techniques to provide an improved estimate of the glottal pulse. Such a HOS based approach should also be more robust to additive Gaussian noise.

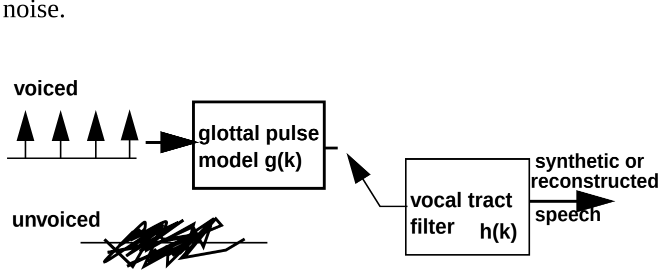

Fig. 2: Speech production model including glottal pulse.

In the work to be presented here, the third-order spectrum or bispectrum will be used. The bispectrum is given by the two-dimensional Fourier transform of the third-order cumulant $c(m, n) = E[x(k)x(k+m)x(k+n)]$:

$$B(\omega_1, \omega_2) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} c(m, n) e^{-j(\omega_1 m + \omega_2 n)} \quad (2)$$

If the signal of interest $x(k)$ is corrupted by additive white Gaussian noise $w(k)$, then $y(k) = x(k) + w(k)$ and from the properties of cumulants [1], [2], we can write

$$c_y(m, n) = c_x(m, n) + c_w(m, n) \quad (3)$$

Taking the bispectrum and noting that for Gaussian noise, all the cumulants of order three and above are identically zero [1], [2], we have that

$$B_y(\omega_1, \omega_2) = B_x(\omega_1, \omega_2) \quad (4)$$

and the bispectrum of $x(k)$ is theoretically free of the influence of noise [1], [2]. Furthermore, as the bispectrum is complex-valued, it retains phase information, allowing the reconstruction of non-minimum phase systems to within a scale factor and a shift (as linear phase is not recoverable) [5], [8].

Noting that the bispectrum has been used successfully with voiced sounds other than vowels [19], [23], the intention is to try bispectrum based approaches with models and real examples of both vowels and other voiced sounds such as nasals or voiced fricatives - which were shown to have significant bispectral features in [9]. The other significant problem wiith the use of the bispectrum is the high variance of estimates. In the work presented here, time domain segment averaging [8] is used to reduce the variance of the bispectrum estimate. As speech is generally regarded as approximately stationary in frames of 10-30ms, a long data record is needed and in this work, where real speech is used, sustained speech sounds from the CD accompanying Childers [28] have been used.

II HOS Factorization and the Linear Bispectrum

In this section, an overview of the theory of HOS factorization and the linear bispectrum is presented. It is then shown how voiced speech can be modelled as a nonGaussian coloured noise driven system, which allows the linear bispectrum theory and the ARMA estimation algorithm described in this section to be applied to estimate the glottal pulse. Some results from simple systems of this type are presented and difficulties that have been uncovered in the use of this technique are discussed.

The bispectrum based approaches to glottal pulse estimation proposed in this paper are based on using the linear bispectrum of signals for the identification of systems driven by a coloured input [12]. The linear bispectrum is that part of the bispectrum which may be modelled by a nonGaussian white noise driven linear time-invariant (LTI) system. Its existence follows from the higher order spectrum factorization theorem [11] applied to systems driven by a coloured input.

According to the higher order spectrum factorization theorem [11], for a bispectrum $B(\omega_1, \omega_2)$, a unique (to a time shift and scaling factor) stable LTI system exists such that

$$B(\omega_1, \omega_2) = \beta H(\omega_1) H(\omega_2) H^*(\omega_1, \omega_2) \quad (5)$$

iff the complex bicepstrum, given by the two-dimensional inverse Fourier transform of the logarithm of the bispectrum

$$c(m_1, m_2) = F^{-2}\{\log[B(\omega_1, \omega_2)]\} \quad (6)$$

exists and is zero everywhere except on the axes.

In [12], it is shown that for a LTI system driven by nonGaussian coloured noise, the bispectrum can be expressed as:

$$B_y(\omega_1, \omega_2) = B^L{}_y(\omega_1, \omega_2) B^D(\omega_1, \omega_2) \quad (7)$$

where $B^L{}_y(\omega_1, \omega_2)$ is called the linear bispectrum and $B^D(\omega_1, \omega_2)$ is that part of the bispectrum of $y(k)$ that is not the linear bispectrum.

It is further shown that this decomposition is unique because, in the bicepstral domain, the complex bicepstra c_y^L and c^D are disjoint [12]. $B^L{}_y(\omega_1, \omega_2)$ is given by

$$B^L{}_y(\omega_1, \omega_2) = B^L{}_x(\omega_1, \omega_2) H(\omega_1) H(\omega_2) H^*(\omega_1, \omega_2) \quad (8)$$

where $B^L{}_x(\omega_1, \omega_2)$ is the linear bispectrum of the input, $x(k)$, and $H(\omega)$ is the system transfer function. It follows from (8) that the finite impulse response (FIR) system model can be extracted from a nonGaussian white noise driven system and from a coloured noise driven system provided that the linear bispectrum of the input process is known [11].

However, in the case that the linear bispectrum of the input process is not known, [12], $B^L{}_y(\omega_1, \omega_2)$ can be modelled as

$$B^L{}_y(\omega_1, \omega_2) = F(\omega_1) F(\omega_2) F^*(\omega_1, \omega_2)$$
$$= \beta \frac{H(\omega_1) H(\omega_2) H^*(\omega_1, \omega_2)}{A(\omega_1) A(\omega_2) A^*(\omega_1, \omega_2)} \quad (9)$$

where the desired system model is the MA part and $B^L{}_x(\omega_1, \omega_2)$ is modelled as the AR part. The inverse of $F(\omega)$ is recovered from the bicepstrum [11] and equations developed from

$$f_{inv}(m) \otimes h(m) = a(m) \quad (10)$$

are solved first to find an estimate for $h(m)$ and then an estimate for $a(m)$, the AR parameters of the linear bispectrum of the input.

A. *Application of the Linear Bispectrum to Glottal Pulse Estimation*

A new model of speech production is shown in Fig. 3. The coloured input $x(k)$ results from the shaping of the nonGaussian white noise input, $e(k)$ (a pseudoperiodic impulse train) by the glottal pulse filter. So from (9) we have

$$B^L{}_y(\omega_1, \omega_2) = \beta_e \frac{V(\omega_1)V(\omega_2)V^*(\omega_1, \omega_2)}{A(\omega_1)A(\omega_2)A^*(\omega_1, \omega_2)} \qquad (11)$$

where β_e is the skewness of the NGWN process $e(k)$, $V(\omega)$ is the system transfer function modelling the vocal tract and $A(\omega) = \frac{1}{G(\omega)}$, the inverse of the system transfer function modelling the glottal pulse.

Applying the algorithm of [12] once provides an FIR estimate for $v(k)$, together with an AR parameter estimate for $g(k)$. What is required ultimately however, is an estimate of the shape of the glottal pulse. Working with a simple model, with a nonGaussian white noise input and simple FIR models for $v(k)$ and $g(k)$ suggested that reapplication of the of the algorithm, once $v(k)$ has been found, would allow $g(k)$ to be determined. Reapplication of the algorithm is achieved by estimating $V(\omega_1)V(\omega_2)V^*(\omega_1, \omega_2)$ and thus finding an estimate of $B^L{}_x(\omega_1, \omega_2)$, the linear bispectrum of $x(k)$, the coloured input process.

Good results were only obtainable with very simple filter models, as illustrated in Fig. 4, which shows results for FIR models $v(k)$ and $g(k)$, obtained from a weighted average of valid estimates, using a 4th order AR estimate in each application of the algorithm of [12]. The results from this approach were only moderately successful and it was not possible to extend this approach to estimation of more complex models. The presence of zeroes on or very near the unit circle in a typical glottal pulse model makes the application of the method difficult as it causes the spectral factorization problem to become ill-conditioned [11].

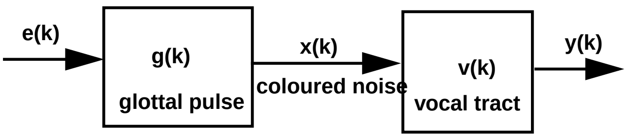

Fig. 3: Application of coloured noise input model to speech production.

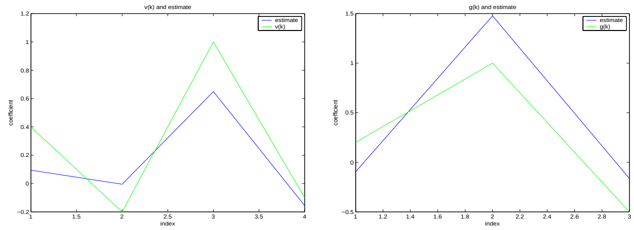

Fig. 4: Model estimation results - 5120 samples, no added noise.

III Application of the Linear Bispectrum to Inverse Filtering

In this section, it is shown how the linear bispectrum approach can be used to obtain alternative glottal pulse and vocal tract estimates in the IAIF approach. This approach is called hybrid IAIF (hIAIF) and results are presented for comparison with results from the IAIF approach. Finally, an innovative joint estimation of the glottal pulse and vocal tract which leads directly to inverse filtering is introduced. This new technique shows promising results and is much simpler than previous iterative techniques as it is a two stage - estimation and inverse filtering - process.

A. *Hybrid Iterative Adaptive Inverse Filtering (hIAIF)*

The most successful inverse filtering algorithm for obtaining the glottal pulse is IAIF [13]. In IAIF, the AR estimates of the glottal pulse and the vocal tract are made by linear prediction using second-order statistics.

hIAIF is illustrated in Fig. 5. As shown in the figure, the AR estimates for the glottal pulse and the vocal tract are made using the linear bispectrum approach based on equations (9) and (10). This is indicated by the notation *LB-n*, where n is the order of the AR estimation, in the figure. At certain points in the process, the linear bispectrum has to be re-calculated.

In the first glottal pulse estimation step, an AR model of order two is estimated, as using a model of order one leads to ill-conditioned matrix equations. The ARMA modelling required by the linear bispectrum approach means that the vocal tract is being simultaneously modelled using an MA model of length 60. In the second glottal pulse estimation step, an AR model of order four is estimated. The vocal tract models estimated are of order 14, while there is a simultaneous MA model of the glottal pulse made. Using synthetic speech models, it has been found, that a length of 10-15 for this MA model seems to allow the simultaneous linear bispectrum estimation to give the best AR estimate of the vocal tract.

The requirement of the bispectrum based approach for a large data record and the time domain segment averaging performed means that it is not pitch synchronous. IAIF was therefore implemented for comparison because it is also not pitch synchronous. Results of both IAIF and hIAIF applied to synthetic speech did not produce very satisfactory results.

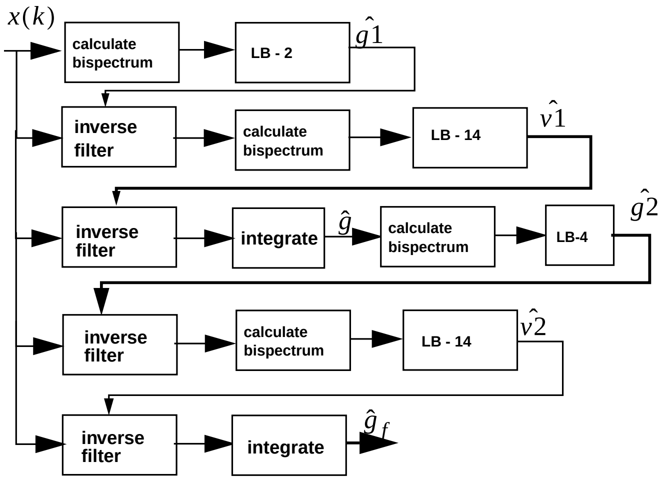

Fig. 5: Hybrid IAIF

The synthetic speech was created using a 128 point Rosenberg glottal pulse [29] and a 14th order vocal tract model. IAIF was applied to a 512 point segment of speech centred on 4 pitch periods. hIAIF was applied to the first 5120 points of the speech. Neither approach coped well with the zero closed phase of the Rosenberg pulse. The hIAIF approach did not remove the ripple due to the first formant.

Tests with real speech sounds show better results. As illustrated in Fig. 6, the two methods were applied to female voiced speech sounds: a vowel sound /aa/ and a voiced fricative /z/. Compared to the IAIF approach, hIAIF does not seem to completely remove the formant ripple. A formant ripple occurs when the glottal contribution is not properly eliminated prior to vocal tract estimation. This effect is worse in the female voice because the higher pitch means that the fundamental frequency is closer to the first formant. In the case of hIAIF, this may be due to the use of order 2 estimation for the initial glottal pulse estimate causing it to include some of the first formant [27]. Improved ARMA estimation techniques could overcome this problem and will be investigated.

It also seems that iteration does not necessarily improve the glottal pulse estimate in hIAIF. This could be due to the higher variance of bispectrum estimates, which could offset the benefits of the iterative process. Another potential disadvantage of hIAIF is the need to recalculate the bispectrum throughout the process, although this can be overcome by utilising techniques implemented in the bicepstral domain which reduce the amount of calculation [12]. On the other hand, the fact that the bispectrum based approach is not necessarily pitch synchronous can be seen as a potential advantage - provided that the results are acceptable.

B. Direct Joint Estimation of Glottal Pulse and Vocal Tract

The disadvantages associated with iteration seem to affect the bispectrum based hIAIF disproportionately, probably due to increased variance in the bispectrum estimates after the various processing steps. As the linear bispectrum approach uses ARMA modelling, it seems reasonable to take advantage of this by jointly modelling the glottal pulse and the vocal tract and then applying inverse filtering only once.

In this approach, based on the vocal tract modelling steps in the hIAIF approach, the vocal tract models estimated are of order 10, while a simultaneous MA model of length 50 is made of the glottal pulse. The AR vocal tract model is used to inverse filter the speech resulting in an estimate of the glottal pulse as shown in Fig. 7.

The result from this approach is very promising and it will be the main strategy to be pursued in further work. The performance of this approach is better with the voiced fricative than with the vowel sound, as was the case with the other bispectrum based techniques presented. A formant ripple is present in the final result for the vowel sound. As discussed above, this effect occurs when the glottal contribution is not properly eliminated prior to vocal tract estimation [27]. However, as in this approach, the glottal pulse and the vocal tract are estimated jointly, it may be more difficult to prevent the closeness of fundamental and first formant causing this problem.

An additional question which needs to be considered is the effect of the strong periodicity of the vowel sound on the bispectrum. In the bispectrum of a vowel sound, this shows up clearly in a distinctly regular patterned appearance, sometimes called the ‘bed of nails’ [9]. The harmonic components of the vowel sound will have consistent phases with respect to each other and as the bispectrum preserves phase information, the effect of the phase of the harmonic components on the phase of the systems estimated (such as a vocal tract model) needs to be investigated [30]. One solution to this problem may be to adopt pitch synchronous analysis as this will ensure that the pitch synchronous frames will have the same phase [9]. This approach will eliminate variation in phase from the harmonic relationships which would otherwise arise when segment averaging on non-pitch synchronous segments is used.

IV CONCLUSIONS

In this paper, a range of techniques involving the application of the bispectrum to estimating the glottal pulse from the speech signal has been presented. In particular, the theory of HOS factorization has been used to model voiced speech as a nonGaussian coloured noise driven system. It has been shown how bispectrum based estimation techniques can be used in a new form of iterative algorithm called hybrid iterative adaptive inverse filtering (hIAIF). Finally, a new simpler inverse filtering approach which utilises joint estimation of vocal tract and glottal pulse from the bispectrum and which has very promising early results has been presented.

V References

[1] C. L. Nikias and M. R. Raghuveer, “Bispectrum Estimation: A Digital Signal Processing Framework,” *Proc. IEEE*, vol. 75, pp. 869-891, 1987.

[2] J. M. Mendel, “Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications,” *Proc. IEEE*, vol 79, pp. 278-305, 1991.

[3] G. Sundaramoothy, M. R. Raghuveer & S A. Dianat, “Bispectral Reconstruction of Signals in Noise: Amplitude Reconstruction Issues,” *IEEE Trans. Acoust., Speech, Signal Processing*”, vol 38, pp. 1297-1306, July 1990.

[4] A. P. Petropulu and C. L. Nikias, “Signal Reconstruction from the Phase of the Bispectrum,” *IEEE Trans. Signal Processing*, vol. 40, no. 3, Mar. 1992.

[5] M. Rangoussi and G.B. Giannakis, “FIR Modeling Using Log-Bispectra: Weighted Least-Squares Algorithms and Performance Analysis,” *IEEE Trans. Circuits Syst.*, vol 38, pp. 281-296, Mar. 1991.

[6] A. P. Petropulu and U. R. Abeyratne, “System Reconstruction from Higher Order Spectra Slices,” *IEEE Trans. Signal Processing*, vol. 45, pp. 2241-2251, Sep. 1997.

[7] M. J. Hinich and M. A. Wolinsky, “A test for aliasing using bispectral components,” *Journal of the American Statistical Association*, vol. 83, pp. 499-502, Jun. 1988.

[8] S. Shahid and J. Walker, "Application of Bispectrum Based Signal Reconstruction to sEMG Signal", in *Proc. Irish Signals and Systems Conf.*, Maynooth, pp. 407-411, June 2001.

[9] J. W. A. Fackrell, *Bispectral Analysis of Speech Signals*, PhD thesis, The University of Edinburgh, 1996.

[10] I. J. Gdoura, P. Loizuo & A. Spanias, “Speech Processing Using Higher Order Statistics,” in *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 1, pp. 160-163, 1993.

[11] A. M. Tekalp and A. T. Erdem, “Higher-Order Spectrum Factorization in One and Two Dimensions with Applications in Signal Modeling and Nonminimum Phase System Identification,” *IEEE Trans. Acoust. Speech Signal Processing*, vol. 37, pp. 1537-1549, 1989.

[12] A. T. Erdem and A. M. Tekalp, “Linear bispectrum of signals and identification of nonminimum phase FIR systems driven by coloured input,” *IEEE Trans. Signal Processing*, vol 40, pp. 1469-1478, 1992.

[13] P. Alku, E. Vilkman and U. K. Laine, “Analysis of glottal waveform in different phonation types using the new IAIF-method,” in *Proc. 12th Internat. Congress Phonetic Sciences*, vol. 4, pp. 362-365, 1991.

[14] T. E. Tremain, “The government standard linear predictive coding algorithm: LPC-10,” Speech *Technol.*, pp. 40-49, Apr. 1982.

[15] A. Gersho, “Advances in Speech and Audio Compression,” *Proc. IEEE*, vol. 82, pp. 900-918, June 1994.

[16] M. Schroeder and B. S. Atal, “Code-excited linear prediction (CELP) high quality speech at very low bit ates,” in *Proc. ICASSP*, pp. 937-940, Mar. 1985.

[17] K. K. Paliwal and M. M. Sondhi, “Recognition of Noisy Speech using Cumulant-based Linear Prediction Analysis,” in *ICASSP’91*, vol. 1, pp. 429-432, 1991.

[18] I. Gerson and M. Jasluk, “Vector Sum Excited Linear Prediction (VSELP) Speech Coding at 8 kbits/s,” in *Proc. ICASSP*, pp. 461-464, Apr. 1990.

[19] M. J. Hinich and E. Shichor, “Bispectral Analysis of Speech” in *Proceedings of the 17th Convention of Electrical and Electronic Engineers in Israel*, pp. 357-360, 1991.

[20] K. S. Lii and M. Rosenblatt, “Deconvolution and estimation of transfer function phase and coefficients for nonGaussian linear processes,” *Annals of Statistics*, vol. 10, pp. 1195-1208, 1982.

[21] D. G. Childers and H. T. Hu, “Speech synthesis by glottal excited linear prediction,” *J. Acoust. Soc. Am.*, vol. 96, pp. 2026-2036, 1994.

[22] H. T. Hu and H. T. Wu, “A Glottal-Excited Linear Prediction (GELP) Model for Low-Bit-Rate Speech Coding,” *Proc. Natl. Sci. Counc. ROC(A)*, vol. 24, pp. 134-142, 2000.

[23] W-T. Chen and C-Y. Chi, “Deconvolution and Vocal-Tract Parameter Estimation of Speech Signals by Higher-Order Statistics based Inverse Filters,” in *Proceedings of IEEE Signal Processing Workshop on Higher Order Statistics*, 1993.

[24] R. A. Wiggins, “Minimum entropy deconvolution,” *Geoexploration*, vol. 16, pp. 21-35, 1978.

[25] G. Fant, J. Liljencrants & Q. Lin, “A four-parameter model of glottal flow,” *STL-QPSR*, vol. 4, pp. 1-13, 1985.

[26] D. Y. Wong, J. D. Markel & A. H. Gray Jr., “Least squares glottal inverse filtering from the acoustic speech waveform,” *IEEE Trans. Acoust. Speech Signal Processing*, vol. 27, pp. 350-355, 1979.

[27] P. Alku, “Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering,” *Speech Communication*, vol. 11, pp. 109-118, 1992.

[28] D. G.Childers, Speech processing and synthesis toolboxes, New York: J. Wiley & Sons, 2000.

[29] A. E. Rosenberg, “Effect of glottal pulse shape on the quality of natural vowels,” *J. Acoust. Soc. Am.*, vol. 84, pp. 583-588, 1971.

[30] G. Kubin, question at Nolisp’03.

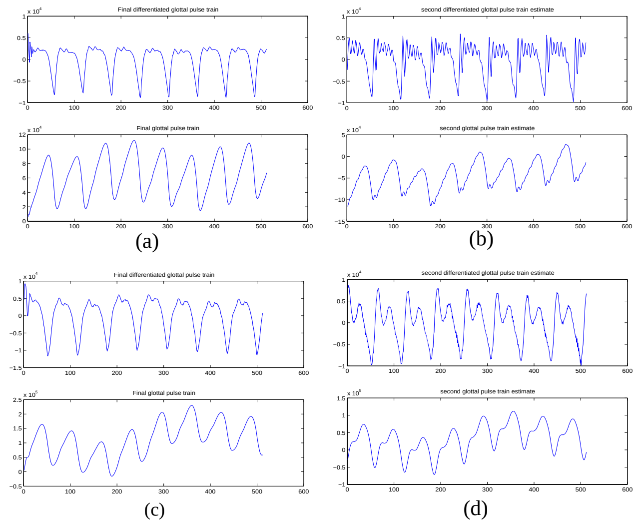

Fig. 6: Glottal pulse recovered from female vowel /aa/ using (a) IAIF and (b) hIAIF and female voiced fricative /z/ using (c) IAIF and (d) hIAIF.

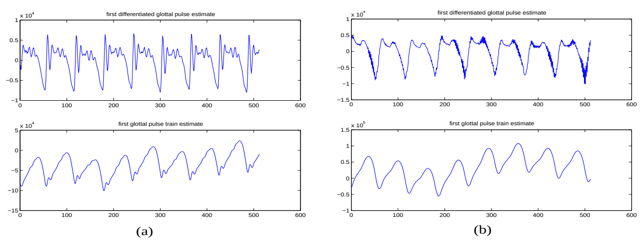

Fig. 7: Joint estimation and inverse filtering: glottal pulse recovered from (a) female vowel /aa/ and (b) voiced fricative /z/.