PITCH DETECTION AND ESTIMATION USING ADAPTIVE IIR COMB FILTERING

Y.K. Jang, J.F. Chicharo and B. Ribbum

Department of Electrical and Computer Engineering
University of Wollongong

ABSTRACT - A modified gradient-based IIR comb filtering technique is presented to estimate and track the pitch period. The advantages of the proposed method compared to the super resolution pitch determination technique (Medan, Yair and Chazan, 1991) are the following: reduced computational burden, improved detection response and the availability of pitch estimates at every sample update Experimental results are included which show the potential performance of the proposed technique.

I. INTRODUCTION

Pitch detection, estimation and tracking continues to be one of the fundamental problems in speech processing. Many different solutions (Dunbnowski Schafer & Rabiner, 1976; Rabiner, 1977; Markel, 1972; Hess, 1983; Medan, Yair & Chazan, 1991) have been devised based on both time and frequency domain approaches. The reality at this current point in time is that no single approach is considered to be both robust and accurate enough for many practical purposes (Hess, 1983; Medan, Yair & Chazan, 1991; Perkis & Ribbum, 1991). In other words, the pursuit for a high performance pitch determination solution is a relevant and current research activity. This is a non trivial task as speech is a highly variable and irregular process from a signal processing point of view. The specific difficulties associated with pitch determination has been discussed in detail by Medan, Yair and Chazan (1991). A recent literature survey reveals that the super pitch method proposed by Medan, Yair & Chazan, (1991) appears to yield the best performance compared to other methods, however; it is computationally expensive.

Recently there has been considerable research activity (Chicharo & Ng, 1990; Chicharo, 1990; Chicharo, 1991) in the area of adaptive notch and comb filtering for the purpose of estimating and tracking sinusoids and harmonic signals in broad band processes. Since pitch can be considered as a periodic process or more precisely a harmonic series, then it is interesting to apply these high performance time domain adaptive techniques to this problem. A direct performance comparison will be carried out between the super resolution pitch determination method and the proposed technique.

The proposed pitch determination system is illustrated in Figure 1 and it uses a modified IIR comb filter structure where only one parameter is adapted which corresponds to the pitch period T_0 of pre-processed signal p_k at time k. The adaptive process is achieved by a simple but exact gradient based algorithm. The speech signal x_k is pre-filtered with a low pass filter as indicated. This low pass filter improves the robustness of the proposed technique as it eliminates some of the local minima in the error surface of the cost function.

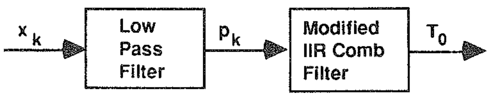


Figure 1. Proposed pitch determination system

This paper is organised as follows: The proposed pitch detection, estimation and tracking system is presented in section II. Section III discusses the performance and compares the results with those obtained using the approach by Medan, Yair & Chazan, (1991). Finally, Section IV concludes the paper.

II. PITCH ESTIMATION SYSTEM

This section first of all discusses a modified comb filter model suitable for a pitch estimator. Secondly, an exact gradient based adaptive algorithm is derived for the proposed comb model structure.

Proposed comb filter model

The proposed comb filter model shown in Figure 2 consists of two blocks. The second block, $H_1(z)$, uses an IIR notch filter parameterized to estimate the fundamental frequency. The first block, $H_m(z)$, is in fact an IIR comb filter which is effectively used to eliminate local minima associated with the frequencies of harmonic components. The local minima problem was reported in Chicharo (1990). Elimination of the local minima ensures that a gradient-based adaptive algorithm converges to the minimum associated with the fundamental component of the harmonic series.

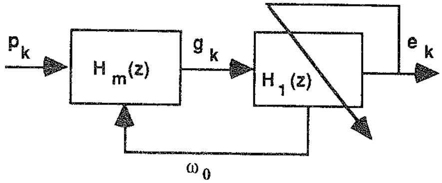


Figure 2. Proposed comb filter model

The notch filter, $H_1(z)$, required to estimate the fundamental frequency is described by the following transfer function:

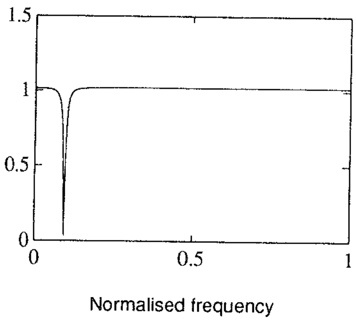
$$H_1(z) = \frac{1 + a_1 z^{-1} + z^{-2}}{1 + \alpha a_1 z^{-1} + \alpha^2 z^{-2}} \tag{1}$$

where $a_1 = -2\cos\omega_0$ and ω_0 is the fundamental angular frequency. Equation (1) was proposed previously in Chicharo (1990) where the parameter α controls the position of the poles and hence the bandwidth of the filter. The parameter α is chosen to be in the range $0 \le \alpha < 1$. The comb filter, $H_m(z)$, for eliminating local minima can also be expressed as the cascaded structure of (m-1) notch filters:

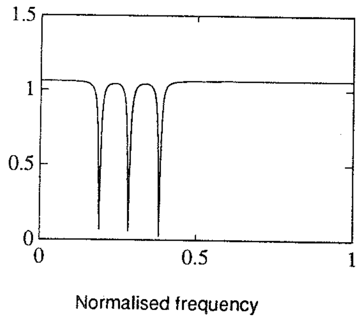
$$H_m(z) = \frac{\prod_{i=2}^{m} (1 + a_i z^{-1} + z^{-2})}{\prod_{i=2}^{m} (1 + \alpha a_i z^{-1} + \alpha^2 z^{-2})} \tag{2}$$

where the filter parameter a_i for i = 2, 3, ······, m is given by

$$a_i = -2\cos i\omega_0 \qquad \text{for} \quad -\pi \le i\omega_0 \le \pi \tag{3}$$



(a) Gain characteristic of notch filter



(b) Gain characteristic of comb filter

Figure 3. Gain characteristics of notch filter $H_1(z)$ and comb filter $H_m(z)$ (ω_0=0.3 and α = 0.98).

The gain characteristics of the notch filter $H_1(z)$ and the comb filter $H_m(z)$ are given in Figure 3. Note that the transfer function $H_1(z)$ characterises a single notch filter structure where the angular frequency ω_0 is estimated which corresponds to the fundamental pitch frequency. The estimate of ω_0 is passed to the comb filter $H_m(z)$ block and tunes the parameters a_i given in Equation (3). Hence the notches of the comb filter in Figure 3 (b) eliminate the local minima associated with integral multiples of the fundamental pitch frequency as the parameter a_1 approaches to the minimum associated with the fundamental pitch frequency. The order of the comb filter (that is, the number of notches) are chosen based on the expected number of higher order harmonics present near the fundamental pitch frequency of the low pass signal. Hence it also depends on the cutoff frequency of the low pass filter.

Gradient based algorithm

Next we derive a gradient based estimation algorithm for the IIR notch filter structure illustrated in Figure 2. Suppose that an input signal g_k is passed through the notch filter model described by Equation (1) at time k. The resulting error signal, e_k, can be expressed as:

$$e_k = (1 + a_1 z^{-1} + z^{-2}) g_k - (\alpha a_1 z^{-1} + \alpha^2 z^{-2}) e_k \tag{4}$$

The notch filter parameter update using a gradient based algorithm (Shynk, 1989) is

$$a_{1,k+1} = a_{1,k} + \mu r_{k+1}^{-1} \nabla_{a_{1,k}} \tag{5}$$

where μ is a factor which controls the rate of convergence. Small values of μ are recommended in order to provide a filtering process which improves the effects of the gradient measurement error (Widrow & Stearn, 1986). The resulting disadvantage is that the convergence rate is reduced. ∇_{a_1} is the partial derivative of the error squared with respect to a_1, that is,

$$\begin{aligned}\nabla_{a_{1,k}} &= \frac{\partial(\frac{1}{2}e_k^2)}{\partial a_{1,k}} = e_k \frac{\partial e_k}{\partial a_{1,k}} \\ &= e_k \left\{ (g_{k-1} - \alpha e_{k-1}) - (\alpha z^{-1} + \alpha^2 z^{-2}) \frac{\partial e_k}{\partial a_{1,k}} \right\}\end{aligned} \tag{6}$$

and r_k is an estimate of the Hessian that is updated by

$$r_{k+1} = (1-\mu) r_k + \mu \left(\frac{\partial e_k}{\partial a_{1,k}} \right)^2 \tag{7}$$

Denoting $\phi_k = \frac{\partial e_k}{\partial a_{1,k}}$ and substituting it into Equation (6) yields

$$\nabla_{a_{1,k}} = e_k \phi_k \tag{8}$$

$$\phi_k = g_{k-1} - \alpha e_{k-1} - \alpha \phi_{k-1} - \alpha^2 \phi_{k-2} \tag{9}$$

The number of operations required to estimate the parameter $a_{1,k}$ are 13 multiplications and 10 additions at each sample update.

The estimate of ω_0 is available at every sample update using this gradient based algorithm. At every sample update the pitch period can be detected and estimated by

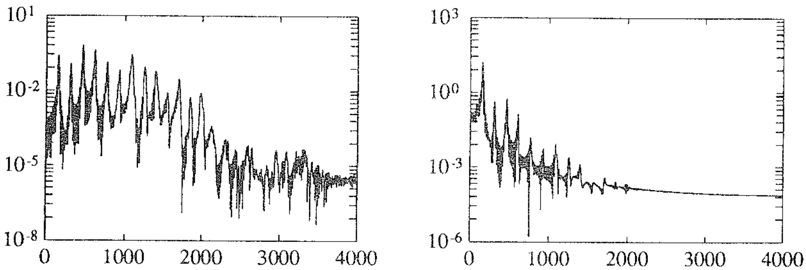
$$T_0 = \frac{2\pi}{\cos^{-1}(-a_{1,k}/2)} \tag{10}$$

The performance of the proposed gradient based algorithm is compared to the performance of the super resolution pitch determination method (Medan, Yair and Chazan, 1991) in the next section. The super resolution pitch determination method is based on measuring the cross correlation in a speech signal, with a frame length equal to the pitch period. The correlation factor is calculated for a suspected pitch period in the range $T_{min} \leq T_0 < T_{max}$. The maximum value of the correlation factor corresponds to the integer pitch period. Hence the exact pitch period is computed by correcting the pitch truncation error. The super resolution pitch determination method provides a pitch estimate at the end of every block of data. For experimental purposes we have chosen a 90 sample block of data. Further details on this approach can be obtained from Medan, Yair and Chazan, (1991).

III. RESULTS AND PERFORMANCE EVALUATION

The proposed algorithm was implemented and tested by applying it to a number of representative speech signals. One of the results is now presented to indicate its performance.

Consider the following example. A data file was constructed by sampling speech at a rate of 8 kHz. This signal was then low pass filtered by passing it through a smooth filter with a cut-off frequency of 100 Hz. A small section of the speech data file signal is shown in Figure 4(a). Figure 4(b) shows the signal after it has been low pass filtered. Clearly the component of fundamental pitch frequency is characterised by the maximum spectral energy after low pass filtering. Hence if we start the adaptive process with a low initial frequency estimate, it is expected that a gradient based algorithm will converge to the fundamental pitch frequency provided μ is small.

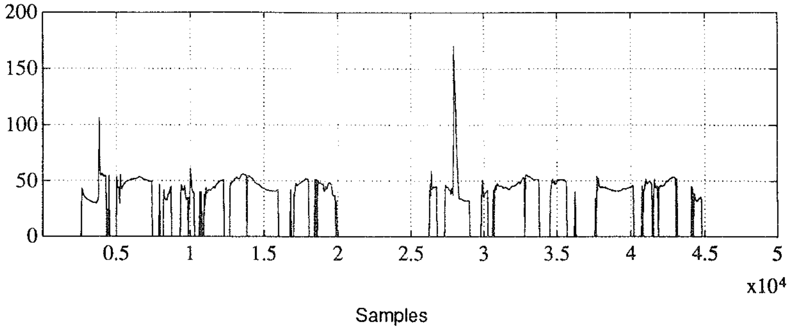


(a) Power spectrum of speech signal (b) Power spectrum of low pass filtered speech signal

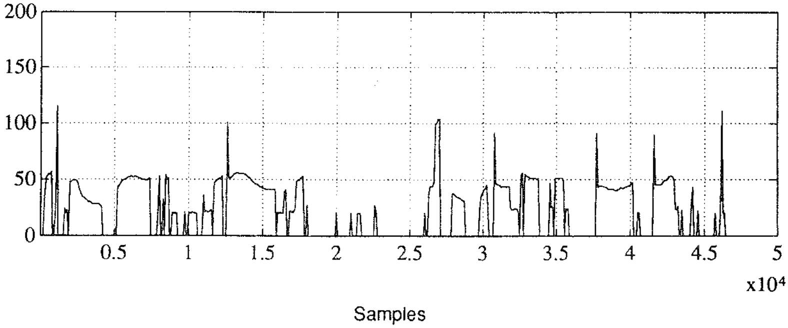
Figure 4. Power spectra of speech signal and filtered signal

The initial estimate of a_1 was chosen to be -1.99 which corresponds to the frequency of 127.2 Hz. The choice of initial estimate is based on the fact that the fundamental frequency is assumed to be close to DC, that is, relatively low frequency. The parameter α controls the bandwidth of the notch filter. The wider the bandwidth the faster the convergence for the gradient based algorithm. The limitation associated with a wider bandwidth is that the resolution of the filter is reduced. In this example α was chosen to be 0.95 which offers a good compromise between resolution and convergence rate performance.

Note that the gradient based algorithm stops its adaptation when no speech signal exists. This is achieved by implementing a simple speech detector based on the signal energy. In other words, if the signal energy (averaged over 10 samples) drops below a certain threshold the adaption process is halted. If subsequently speech is again detected the system is reset and the previously determined pitch period is used as an initial estimate. Figure 5 (a) shows the results of the pitch period estimates using the gradient based adaptive algorithm. Clearly the proposed scheme is capable of providing reliable pitch estimates provided the speech detector is successful.



(a) Pitch period using IIR comb filtering



(b) Pitch period using super resolution pitch determination technique

Figure 5. Pitch period of fluent speech

Figure 5 (b) shows the results obtained using the super resolution pitch method by Medan, Yair and Chazan, (1991) on the same speech data file. It is evident that the results of the proposed technique compare very well with the results obtained from the super resolution pitch determination method but at a much reduced computational burden. The number of multiplications required to estimate 90 pitch periods was approximately 2000 using the proposed adaptive gradient method. Using the super resolution pitch determination method, the number of multiplications per block (the block length = 90) was approximately 9500 when T_{min} =20 and T_{max} =135. The lower limit for the pitch search, T_{min} corresponds to a 400 Hz and the upper limit T_{max} is half the number of the buffered samples. As stated previously the proposed method provides pitch estimates at every sample update which means that it is capable of faster estimation response. The exact improvement in terms of response is difficult to establish as it depends on many factors including the adaption coefficient μ. However, based on the limited number of trials conducted so far it appears that the response is better than the super pitch method.

To evaluate the pitch estimation accuracy consider the example shown in Figure 6. The estimate of pitch period using the gradient based algorithm was found to be as good or better than the super resolution pitch determination method. This was confirmed by spectral analysis of the specific areas in the data file where a discrepancy occurred between the two methods. Note that at this stage further and more comprehensive analysis is required.

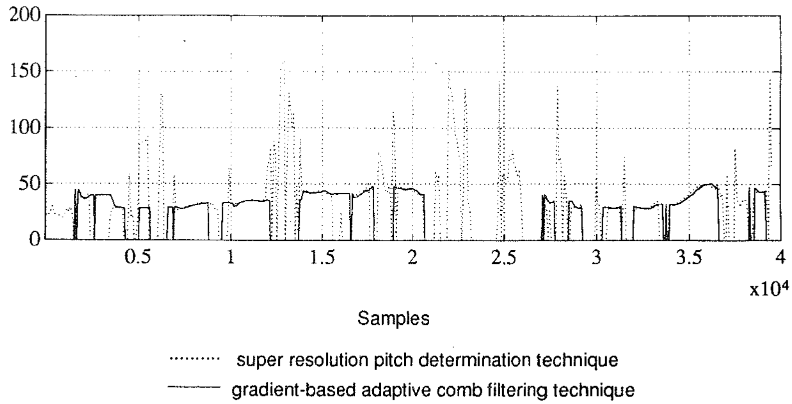


Figure 6. Pitch period in number of samples

IV. CONCLUSION

This paper has presented a novel approach for the task of pitch estimation. The method is based on a gradient based adaptive comb filter structure. The main difficulty of an adaptive filtering approach for the pitch problem is primarily the existence of local minima in the adaption processes. We have resolved this issue by firstly low-pass filtering the speech signal and secondly using a series of notches which are harmonically related to eliminate any further existing local minima. The results obtained were compared to the super pitch method recently published by Medan, Yair and Chazan, (1991). Based on preliminary studies the proposed approach provides pitch estimates which are as good or better as far as accuracy is concerned when compared to the super pitch approach. In addition the proposed method is superior in terms of being characterised by a much reduced computational burden as well as providing pitch estimates at every sample update rather than at the end of every block of data. It should noted however, that further and more comprehensive analysis and verification is required.

V. REFERENCES

Chicharo J.F. & Ng. T.S. (1990) *Gradient-based IIR notch filtering for frequency estimation*, IEEE Trans. Acoust. Speech and Signal Processing, ASSP-38, 769-777.

Chicharo J.F. (1990) *Error surface analysis of adaptive comb filter*, Electronic Letters, 26, 587-589.

Chicharo J.F. (1991) *High resolution spectral estimation using specially constrained adaptive notch filter*, Int. J. Electronics, 72 57-66.

Dunbnowski J.J., Schafer R.W. & Rabiner L.R. (1976) *Real time digital hardware pitch detector*, IEEE Trans. Acoust. Speech and Signal Processing, ASSP-24 2-8.

Hess W. (1983) *Pitch determination of speech signals* (Springer: New York).

Markel J.D. (1972) *The SIFT algorithm for fundamental frequency estimation*, IEEE Trans. Audio Electroacoust., AU-20, 367-377.

Medan Y., Yair E. & Chazan D. (1991) *Super resolution pitch determination of speech signals*, IEEE Trans. on Signal Processing, Vol 39, 40-48.

Perkis A. & Ribbum B. (1991) *Application of stochastic coding schemes in satellite communication*, Advances in Speech Coding, edited by: Atal, Cuperman and Gersho, Kluwer Academic Publishers.

Rabiner L.R. (1977) *On the use of autocorrelation analysis for pitch determination*, IEEE Trans. Acoust. Speech and Signal Processing, ASSP-25, 22-33.

Shynk J.J. (1989) *Adaptive IIR filtering*, IEEE ASSP Magazine.

Widrow B. and Stearn S.D. (1986) *Adaptive signal processing* (Prentice-Hall : NJ).