NOISE-ROBUST PITCH DETECTION BASED ON BIORTHOGONAL WAVELET TRANSFORM AND FORMANT-FREQUENCY TRACING ALGORITHM

Shi-Huang Chen and Jhing-Fa Wang

Department of Electrical Engineering,
National Cheng Kung University, Tainan, Taiwan 701, R.O.C.
Email: shchen@cad.ee.ncku.edu.tw & wangjf@server2.iie.ncku.edu.tw

ABSTRACT

In this paper, an improved wavelet-based pitch determination method is presented. Based on the biorthogonal wavelet transforms and a new formant-frequency tracing algorithm, the proposed method can accurately and efficiently extract the pitch information from the speech signal even for noisy environments. In this new method, the input speech will be first decomposed into a formant-contain approximation signal using the biorthogonal wavelet transforms and the formant-frequency tracing algorithm. Then, the pitch information can be detected in the formant-contain approximation signal via an object-depended maximum extractor and a pitch correction algorithm. Besides the robustness against noise, the proposed method does not embed any segmentation window and predefined threshold which are usually needed in the conventional wavelet-based pitch detection algorithms. From the experimental results, one can find that the proposed method works well for synthetic and natural speech signals. In addition, it is also shown that the proposed algorithm can achieve the better performance than those of the present wavelet-based as well as the conventional time-domain and spectral-domain pitch detectors under noisy conditions.

1. INTRODUCTION

The extraction of pitch information is frequently regarded as an essential task in a variety of speech processing applications. Here, the pitch information is referred to both the pitch period and the instants of glottal closure (GCI's) in voiced speech. The pitch detectors can be applied to the high-quality speech coding system [1], the PSOLA based text-to-speech (TTS) system [2], and the speaker verification system [3]. Although many pitch detection algorithms including in the spectral, time or wavelet domains have been proposed in the literatures [1-6], the accurate and robust speech pitch determination still remains an open problem.

The conventional pitch detectors, such as spectral-domain and time-domain methods, usually have to use certain thresholds in their algorithms. It is well known that inappropriate selection of the thresholds, regardless of the input speech characteristics, will result in performance degradation. Therefore, they are not suitable for wide range of speakers. In addition, most of them use some kind of the segmentation window to determine the pitch information in a small interval. However, the size and shape of the segmentation window as well as the position of segmentation window with respect to the speech can affect the estimated spectral characteristics. It was also shown that the fundamental frequency significantly influences the spectrum if the window length is longer than one pitch period [7]. Hence, there will be some ambiguities in the locations of the GCI's. Moreover, the performance of these methods generally deteriorates when the speech signal is corrupted by noise and distortion [7].

In [5], a wavelet-based algorithm was proposed for the determination of pitch information. This method was based on the Mallat's work on image [8] essentially. It was assumed that the locations of GCI's in the original speech waveform will the same as those in the several consecutive scales of the wavelet representations. Therefore, the pitch information can be extracted from these wavelet representations instead of original speech waveform. Most of the subsequent wavelet-based pitch detection algorithms are originally inspired by the work presented in [5]. On account of the multiresolution analysis and the time-frequency localization characteristics of wavelet transforms, the pitch detectors embedded wavelet transforms, in fact, have the better performances than those of the conventional types [5]. However, these present wavelet-based methods still have some drawbacks. First, the segmentation windows and the predetermined thresholds are still used in certain of the wavelet-based pitch detection methods [9-10]. As mentioned previously, the performance of these methods will critically depend on the segmentation windows and the predetermined thresholds. Secondly, the wavelet bases applied in most of these present wavelet-based pitch detection algorithms are orthogonal bases. From the viewpoint of robustness, the main disadvantage of using the orthogonal basis is that the white Gaussian noise remains white after orthogonal transforms [11]. Above phenomena also explain why the orthogonal wavelet-based pitch detectors do not have satisfied results in some kinds of noisy conditions.

To overcome the drawbacks described above, this paper develops an improved wavelet-based pitch detection algorithm. Based on the biorthogonal wavelet transform, the input speech signal will be first decomposed into a formant-contain approximation signal whose decomposition level is determined by a new formant-frequency tracing algorithm. Then, the locations of GCI's can be effortlessly detected in the formant-contain approximation signal via an object-depended maximum extractor. Finally the pitch correction algorithm is exploited to achieve an accurate determination of GCI's and the pitch period. It is worthwhile to note that the segmentation windows and the predefined thresholds are complete avoided in this new type of algorithm. To illustrate this, the proposed method is applied to both the synthetic and the natural speech signals. It yields a considerable performance improvement compared with other conventional methods and wavelet-based methods under noisy conditions.

The remainder of this paper is organized as follows. In Section 2, the bases of the biorthogonal wavelet transform and filter bank are described briefly. In Section 3, the detailed description of the proposed pitch detection algorithm will be given. In Section 4, the various experimental results are illustrated. Section 5 concludes the paper.

2. BIORTHOGONAL WAVELET TRANSFORM AND FILTERBANK

Since the wavelet transform provides a tool for both the time-frequency localization and the multiresolution analysis, it was shown that the wavelet is highly efficient in representing the non-stationary signals such as speech [12]. Theoretically, depended on the definitions of orthogonality of the wavelet bases, the wavelet transforms can be categorized into two types, named the orthogonal and the biorthogonal wavelets [13-14]. The main difference between the orthogonal and the biorthogonal wavelet transform is the latter one has two wavelets, one (denoted $\tilde{\psi}(t)$) for decomposition and the another one (denoted $\psi(t)$) for reconstruction. The wavelet $\psi(t)$ and the its dual wavelet $\tilde{\psi}(t)$ can be generated from two companions $\phi(t)$ and $\tilde{\phi}(t)$, which are known as the scaling and the corresponding dual scaling function. Above four scaling and wavelet functions have to satisfy the following two-scale difference equations: [13-14]

$$\phi(t) = \sqrt{2}\sum_n h(n)\phi(2t-n) \ , \tag{1}$$

$$\tilde{\phi}(t) = \sqrt{2}\sum_n \tilde{h}(n)\tilde{\phi}(2t-n) \ , \tag{2}$$

$$\psi(t) = \sqrt{2}\sum_n g(n)\phi(2t-n) \ , \tag{3}$$

$$\tilde{\psi}(t) = \sqrt{2}\sum_n \tilde{g}(n)\tilde{\phi}(2t-n) \ . \tag{4}$$

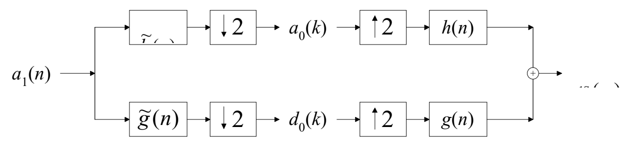


Fig. 1. Two-channel biorthogonal analysis/synthesis filter bank

These four dilation equations listed above can be related to the two-channel biorthogonal filter bank shown in Fig. 1 where $h(n)$ in Eq.(1) and $\tilde{h}(n)$ in Eq. (2) are low-pass filters, whereas g(n) in Eq. (3) and $\tilde{g}(n)$ in Eq. (4) are high-pass filters. Also, the symbols ↓2 and ↑2 shown in Fig. 1 denote the downsampling by 2 and the upsampling by 2, respectively. It is shown [13-14] that to perform the biorthogonal wavelet transform, one therefore dose not require the explicit forms of $\phi(t)$, $\tilde{\phi}(t)$, $\psi(t)$ and $\tilde{\psi}(t)$ but only depends on $\tilde{h}(n)$, $\tilde{g}(n)$, $h(n)$, and $g(n)$. Let $\{a_1(n)\}_{n\in Z}$ be the input to the analysis filter bank as shown in Fig. 1. Then the outputs of the analysis filter bank are given by

$$a_0(k) = \sum_n \tilde{h}(n-2k)a_1(n) \tag{5}$$

$$d_0(k) = \sum_n \tilde{g}(n-2k)a_1(n) \tag{6}$$

where $a_0(k)$ and $d_0(k)$ are now called the approximate coefficients and the detail coefficients of the first level biorthogonal wavelet transform of $a_1(n)$, respectively. The output of the synthesis filter bank shown in Fig. 1 is

$$a_{1S}(m) = \sum_{k\in Z} a_0(k)h(m-2k) + \sum_{k\in Z} d_0(k)g(m-2k) \ . \tag{7}$$

For perfect reconstruction, i.e., $a_1(n) = a_{1S}(m)$, these four filters have to be related as [13]

$$\tilde{g}(n) = (-1)^n h(1-n), \quad g(n) = (-1)^n \tilde{h}(1-n) \ . \tag{8}$$

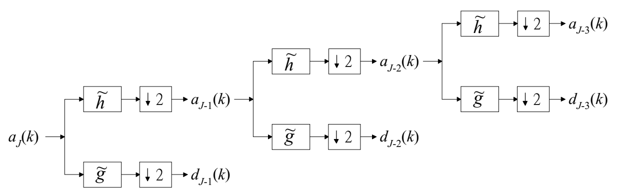


Fig. 2. 3-level two-channel biorthogonal analysis filter banks

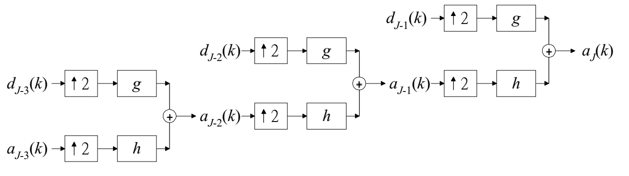


Fig. 3. 3-level two-channel biorthogonal synthesis filter banks

Note that the Eqs.(5) and (6) provide an efficient pyramidal algorithm that uses $\tilde{h}(n)$ and $\tilde{g}(n)$ for the biorthogonal wavelet decomposition. Each approximate coefficients can similarly go through the next level of decomposition. An illustrative example of 3-level two-channel biorthogonal analysis filter bank is given in Fig. 2

and its corresponding synthesis filter bank is given in Fig. 3.

3. THE PROPOSED PITCH DETECTION ALGORITHM

This section gives the detailed discussions about the proposed pitch detection algorithm. First, it will describe the formant-frequency tracing algorithm which is used to obtain the formant-contain approximation signal. Then, it will give a method to extract the pitch period and the GCI's from the formant-contain approximation signal by using an object-depended maximum extractor and a pitch correction algorithm. Finally the investigation of selecting the biorthogonal wavelet basis for pitch detection will be given.

3.1 Formant-Frequency Tracing Algorithm

The objective of the formant-frequency tracing algorithm is to generate a formant-contain approximation signal from the wavelet transform of input speech signal. In this algorithm, the formant-contain approximation signal is defined as a wavelet representation signal that contains most of the first formant frequency elements in input speech signal. Since the first formant frequency directly relates to the pitch period, the proposed algorithm therefore can effortlessly extract the pitch information from the formant-contain approximation signal. In order to accomplish this objective, the decomposition level of wavelet transform is needed to determine appropriately. In fact, this determination procedure should be regarded as the most important step in all the wavelet-based pitch detection algorithms. However, due to the first formant frequencies of human speech signals can vary from 25Hz to 1000Hz [5], most of the present wavelet-based pitch detection algorithms use a predetermined decomposition level to generate a fixed-bandwidth wavelet representation signal. Generally, the bandwidth of this signal is below 1000Hz [5, 9-10].

The main disadvantage of above setting is that the GCI is not the only one whose peak value is the local maximum in the wavelet representation signal. There may exist other harmonic components which are coexisted with the first formant frequency components. It goes without saying that these harmonic components will affect the performance of pitch detection. Consequently, these present wavelet-based pitch detection methods have to utilize some kind of segmentation windows and predetermined thresholds to eliminate the effects of these harmonic components. As mentioned in the preceding section, this way does not meet the requirements of robustness and accurateness.

The proposed formant-frequency tracing algorithm is described as follows. Let $\{a_1(n)\}_{n\in Z}$ be the input speech signal and Fs be its Nyquist frequency. If $J\in N$ is the dominant decomposition level, than the following equation must be satisfied.

$$E\left\{d_{1-(J+m)}(k)\right\} > E\left\{a_{1-(J+m)}(k)\right\},\ m\in N$$

$$\text{and}\quad (J+m) < \left\lfloor \log_2\left(\frac{Fs}{20\text{Hz}}\right)\right\rfloor \tag{9}$$

where $a_{1-(J+m)}(k)$ and $d_{1-(J+m)}(k)$ are the approximate coefficients and the detail coefficients of the $(J+m)$th level biorthogonal wavelet transform of $a_1(n)$. And $E\{f(k)\}$ is denoted the energy of $f(k)$ and is given by

$$E\{f(k)\} = \sum_k |f(k)| \,. \tag{10}$$

Then, one can obtain the formant-contain approximation signal $A_J(t)$ via

$$A_J(t) = \sum_{0\le j\le J-1}\sum_{k\in \mathbf{Z}} a_{1-j}(k)2^j\phi(2^j t-k) \,. \tag{11}$$

3.2 Extract the Pitch Information from Formant-Contain Approx. Signal

Based on the multiresolution analysis of wavelet transforms, the GCI is also marked by a sharp discontinuity in the formant-contain approximation signal $A_J(t)$. And it can be detected in either the positive part of $A_J(t)$ or the negative part of $A_J(t)$ due to the symmetric oscillation characteristics of speech waveform. The positive part of $A_J(t)$ is denoted $PA_J(t)$ and is defined as

$$PA_J(t) = \frac{1}{2}\left[A_J(t) + |A_J(t)|\right] \,. \tag{12}$$

Similarly, the negative part of $A_J(t)$ is denoted $NA_J(t)$ and is defined as

$$NA_J(t) = \frac{1}{2}\left[-A_J(t) + |A_J(t)|\right] \,. \tag{13}$$

The proposed algorithm will select one of them whose variance is small than that of another one for determining the locations of GCI's. Once the $PA_J(t)$ or $NA_J(t)$ is selected, it will be smoothed using a Gaussian filter to enhance the GCI's and to remove the discontinuous in $PA_J(t)$ or $NA_J(t)$ before extracting pitch information. Let $B(t)$ be the selected $PA_J(t)$ or $NA_J(t)$, then the above filtering operation can be expressed as

$$\hat{B}(t) = B(t) * G(t) \tag{14}$$

where * denotes the convolution operation, and $G(t)$ is the Gaussian function defined as

$$G(t) = \frac{1}{(\sigma^2\pi)^{1/4}} e^{-\frac{t^2}{2\sigma^2}} \tag{15}$$

with σ^2 being the variance of $B(t)$ and its duration is 100 sampling points.

Then the GCI detection in $\hat{B}(t)$ can be accomplished via an object-depended maximum extractor. It means that if $t = t_0$ is a GCI in $\hat{B}(t)$, than the following relations must be satisfied.

$$\hat{B}(t_0) > \hat{B}(t_0 - 1) > \cdots > \hat{B}(t_0 - P) \quad \text{and} \quad (16a)$$

$$\hat{B}(t_0) > \hat{B}(t_0 + 1) > \cdots > \hat{B}(t_0 + P) \quad 16b)$$

where P is defined as

$$P = \left\lfloor \frac{\text{The duration of } G(t)}{\text{Variance of } \hat{B}(t)} \right\rfloor . \quad (17)$$

Finally, the accurate locations of GCI's and the pitch period of input speech signal are obtained by a pitch correction algorithm as follows.

(1) Calculate the average distance Da between two adjacent GCI's.

(2) Eliminate the GCI whose distance between its adjacent GCI's is shorter than $0.5Da$ or longer than $2Da$.

(3) Repeat (1) and (2) until no unsuitable GCI is available.

(4) Locate the final existed GCIs' positions and calculate the reciprocal of average distance between two adjacent GCI's as the estimated pitch period.

3.3 The Selection of Biorthogonal Wavelet Basis

Due to the spline functions are symmetric, adjustable, and have FIR filter coefficients, the biorthogonal wavelet systems chosen in the proposed algorithm are biorthogonal spline wavelets developed in [14]. In this type of biorthogonal wavelet functions, the $\tilde{\phi}(t)$ and $\tilde{\psi}(t)$ are both spline functions of compact support [14]. In addition, to be valid for the robust pitch detection required in this paper, the scaling function $\tilde{\phi}(t)$ and its corresponding analysis lowpass filter $\tilde{h}(n)$ have to satisfy the following criteria:

(1) The scaling function $\tilde{\phi}(t)$ should correspond to a reasonably sharp function in order to match the shape of GCI.

(2) The frequency response of $\tilde{h}(n)$ can emphasize certain frequency channels in order to increase the robustness against noise in other frequency channels.

Motivated by the first criterion, four members of the biorthogonal spline wavelets, called B(2.4), B(2.6), B(2.8) and B(6.8), are selected and their waveforms are plotted in Fig. 4. One can find that these analysis scaling functions have the similar shapes, however, the frequency responses of their corresponding analysis lowpass filter $\tilde{h}(n)$ are quite different. The frequency responses of analysis lowpass filters with B(2.4), B(2.6), B(2.8), and B(6.8) are shown in Fig. 5. From the viewpoint of the second criterion, the B(2.8) is the most appropriate biorthogonal spline wavelet for the proposed algorithm. For more detailed information about the biorthogonal spline wavelets, see [13].

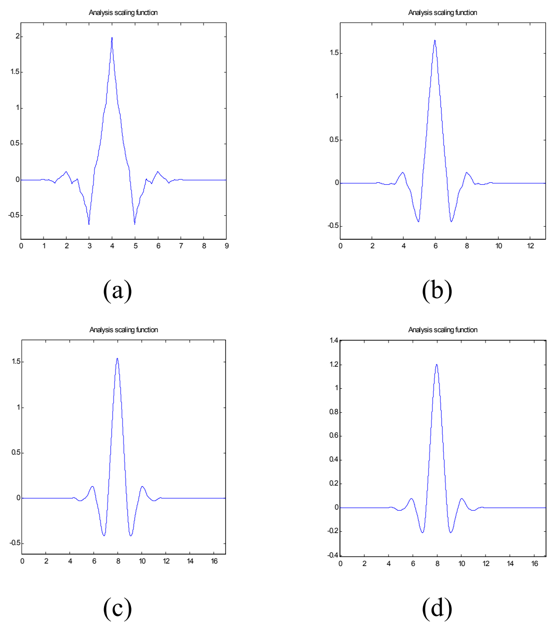


Fig. 4. The $\tilde{\phi}(t)$ of (a) B(2.4), (b) B(2.6), (c) B(2.8), (d) B(6.8).

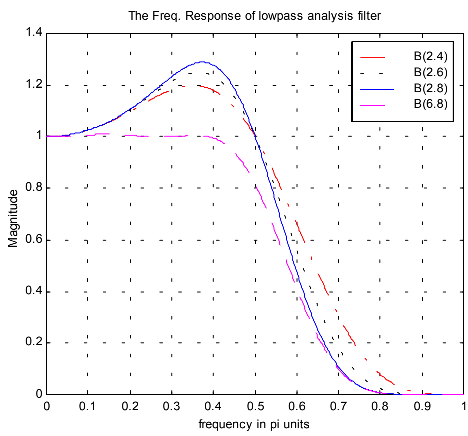


Fig. 5. The frequency response of analysis lowpass filters.

4. PERFORMANCE EVALUATION AND EXPERIMENTAL RESULTS

In this section, it will first discuss the performance of the proposed method on synthetic and natural speech data. The robustness of proposed method is then examined for additive white Gaussian noise. In all the illustrations to follow, the speech signals were sampled at 8 kHz with 8-bit resolution.

4.1 Illustration of the Method for Synthetic Speech Data

The synthetic speech data used in this paper consist of five kinds of voiced phonemes, namely, /a/, /e/, /i/, /o/ and /u/.

Fig. 6(a) shows a synthetic speech signal /a/ whose pitch period is constant and equal to 10ms, and Fig. 6(b) is its formant-contain approximation signal (decomposition level = 5). Fig 6(c)-(d) show the waveform of its corresponding $\hat{B}(t)$ and the locations of GCI's derived from $\hat{B}(t)$ in (c), respectively. The estimated pitch period is 9.9ms.

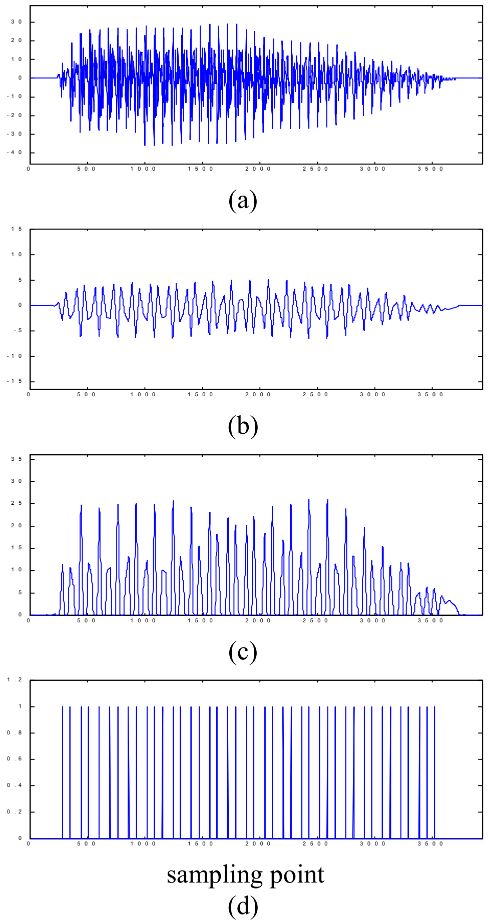


Fig. 6. (a) Clean synthetic speech signal /a/. (b) The corresponding formant-contain approximation signal. (c) Waveform of $\hat{B}(t)$ derived from the signal in (b). (d) The locations of GCI's derived from the signal in (c).

Table 1. The experimental results of the synthetic speech data

Pitch Period / Speech Data	5 ms	10 ms	15 ms	20 ms	25 ms
	Estimated pitch period (ms)				
/a/	4.90	9.90	14.90	19.84	24.90
/e/	4.89	9.86	14.80	19.90	24.85
/i/	4.91	9.88	14.85	19.82	24.82
/o/	5.00	9.92	14.80	19.80	24.80
/u/	4.90	9.90	14.90	19.82	24.92
Average *Err*	1.6 %	1.0 %	1.0%	0.82%	0.57%

Table 1 gives the experimental results of this new method on other synthetic speech data. The error rate (denoted *Err*) used in Table 1 is defined as

$$Err = \frac{|\hat{T} - T|}{T} \times 100\% \qquad (18)$$

where $\hat{T}$ is the estimated pitch period and the T is the true pitch period of the tested synthetic speech signal. The average error rate of the proposed method on detecting the pitch period of synthetic speech signal is below 1.5%.

4.2 Illustration of the Method for Natural Speech Data

Fig. 7 illustrates the experimental result of proposed method on the syllable /chii/ spoken by a male voice. This syllable consists of a unvoiced segment /ch/ and a voiced segment /i/. Fig. 7(a)-(d) show the original speech waveform, the corresponding formant-contain approximation signal (decomposition level = 5), the corresponding waveform of $\hat{B}(t)$, and the estimated locations of GCI's, respectively.

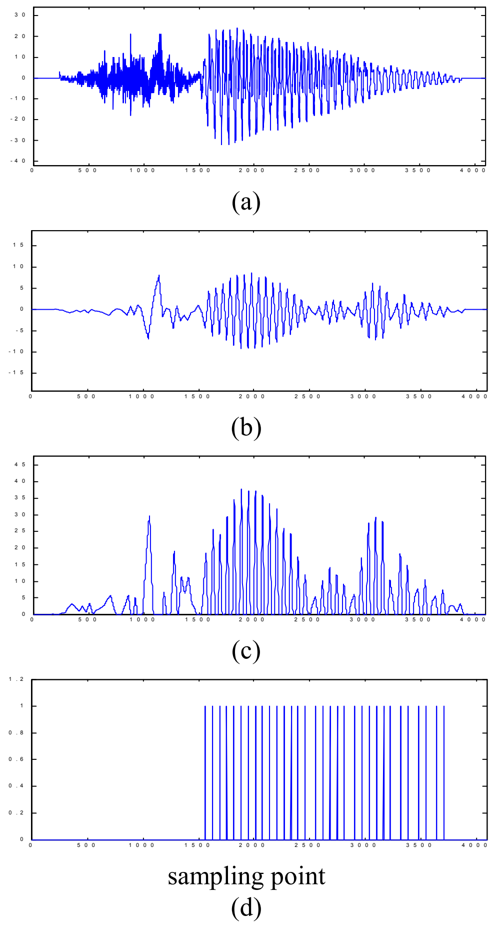
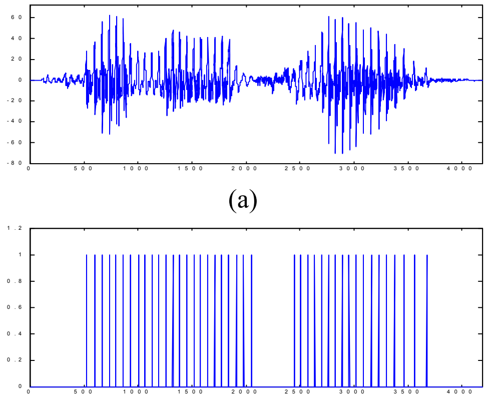


Fig. 7 (a)Clean natural speech signal /chii/. (b) The corresponding formant-contain approximation signal. (c) Waveform of $\hat{B}(t)$ derived from the signal in (b). (d) The locations of GCI's derived from the signal in (c).

In the analysis of continuous speech signal, the each syllable in the input speech sentence will be separated by the segmentation method given in [15]. And then the pitch information of each syllable can be extracted individually. Fig. 8 illustrates the experimental result of proposed method on the initial part of the utterance "ANY TIME any where…." Spoken by a male voice. Fig. 8(a) and (b) show the speech waveform and the estimated locations of GCI's, respectively.



sampling point
(b)

Fig. 8. Illustration of the method for a continuous speech "ANY TIME…": (a) Speech waveform; (b) the locations of GCI's.

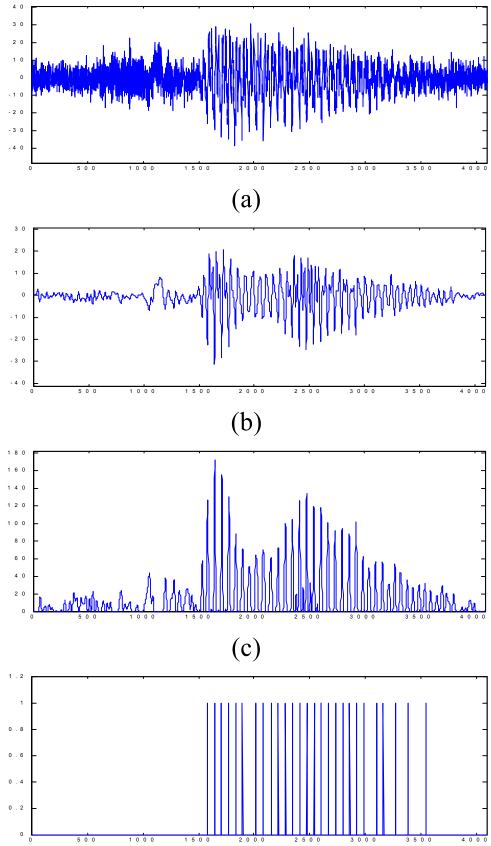
4.3 Robustness of the Method

Finally, the robustness of the proposed method is evaluated under additive noisy conditions. A white Gaussian noise was added to the clean speech, and the performance is evaluated at SNR's 30, 25, 20, 15, 10, 5, and 0dB, respectively. To illustrate this, Fig. 9 shows the experimental results of the natural speech signal of Fig. 7(a) at a SNR of 5dB.

Table 2. Performance of the proposed algorithm at different SNR's

SNR (dB) / Methods	30	25	20	15	10	5	0
	Average *Err* (%)						
Wavelet-based in [5]	1.2	1.4	2.7	4.9	7.5	12.4	23.6
Wavelet-based in [10]	1.1	1.3	2.4	4.2	6.9	8.6	17.8
Cepstrum-based in [6]	1.4	1.6	3.2	5.7	8.6	16.8	32.2
Time-domain in [1]	0.7	1.2	2.6	5.3	10.4	28.7	49.6
Proposed method	1.1	1.2	2.2	3.4	5.1	7.9	12.1

The robustness of the proposed method is also compared against the other present pitch detection methods including in the spectral, time and wavelet domains, and the results are given in Table 2. The definition of *Err* used in Table 2 is similar to that used in Table 1 with modifying $\hat{T}$ be the estimated pitch period under noisy condition and the T be the pitch period of clean speech. From these experimental results, one can find that the pitch information can be accurately and effortlessly extracted by using the proposed method.



sampling point
(d)

Fig. 9. (a) Natural speech of Fig. 7(a) with 5dB Gaussian white noise. (b) The corresponding formant-contain approximation signal. (c) Waveform of $\hat{B}(t)$ derived from the signal in (b). (d) The locations of GCI's derived from the signal in (c).

5. CONCLUSIONS

An improved wavelet-based pitch detection algorithm was presented. In this paper, the biorthogonal spline wavelet transform and the formant-frequency tracing algorithm were proposed to improve the accuracy and the robustness of the conventional wavelet-based pitch detection algorithm. In comparison with the other present pitch detection schemes, the proposed algorithm does not need to use any segmentation window as well as the predetermined threshold. The performance of the proposed algorithm was evaluated on synthetic and natural speech signals. Compared to other time, spectral and wavelet domain pitch detection methods, it was also shown that the proposed method has the better performance under noisy conditions.

6. ACKNOWLEDGMENT

This work was supported by the National Science Council (NSC), Taiwan, R.O.C., under Grant NSC89-2213-E-006-077.

7. REFERENCES

[1] A. M. Kondoz, *Digital Speech Coding for Low Bit Rate Communications Systems*, West Sussex, England, John Wiley & Sons Ltd, 1994.

[2] Fàbio Violaro and Olivier Böeffard, "A Hybrid Model for Text-to-Speech Synthesis," IEEE Trans. on Speech and Audio Processing, Vol. 6, No. 5, pp. 426-434, Sept. 1998.

[3] R. Rabiner and R. W. Schafer, Digital Processing of Speech Signals. Englewood Cliffs, NJ: Prentice-Hall, 1978.

[4] F. Wang, C. H. Wu, S. H. Chang, and J. Y. Lee, "A Hierarchical Neural Network Model Based on a C/V Segmentation Algorithm for Isolated Mandarin Speech Recognition," IEEE Trans. on Signal Processing, Vol. 39, No. 9, pp. 2141-2146, September 1991.

[5] Kadambe and G. Faye Boudreaux-Bartels, "Application of the Wavelet Transform for Pitch Detection of Speech signals," IEEE Trans. on Information Theory, Vol. 38, No. 2, pp. 917-924, March 1992.

[6] Sassan Ahmadi and Andreas S. Spanias, "Cepstrum-Based Pitch Detection Using a New Statistical V/UV Classification Algorithm," IEEE Trans. on Speech and Audio Processing, Vol. 7, No. 3, pp. 333-338, May 1999.

[7] B. Yegnanarayana, and R. N. J. Veldhuis, "Extraction of Vocal-Tract System Characteristics from Speech Signals," IEEE Trans. On Speech and Audio Processing, Vol. 6, No. 4, pp. 313-327, July 1998,

[8] G. Mallat and S. Zhong, "Characterization of Signals from Multiscale Edges," IEEE Trans. of Patt. Analy. and Mach. Intell., Vol. 14, pp. 710-732, July 1992.

[9] Christopher Wendt and Athina P. Petropulu, "Pitch Determination and Speech Segmentation Using the Discrete Wavelet Transform," ISCAS 96, Atlanta, GA, USA, pp. 45-48, 1996.

[10] Shi-Huang Chen and Jhing-Fa Wang, "A Pyramid-Structured Wavelet Algorithm for Detecting Pitch Period of Speech Signal," 1998 International Computer Symposium (ICS), pp. 50-56, Dec. 1998.

[11] Sidney Burrus, Ramesh A. Gopinath and Haitao Guo, Introduction to Wavelets and Wavelet Transforms. Upper Saddle River, NJ: Prentice-Hall, 1998.

[12] Jhing-Fa Wang, Shi-Huang Chen and Jyh-Shing Shyuu, "Wavelet Transforms for Speech Signal Processing," Journal of The Chinese Institute of Engineers, Vol. 22, No. 5, pp.549-560, Sept. 1999.

[13] Daubechies, Ten Lectures on Wavelets, CBMS, SIAM publ., 1992.

[14] A. Cohen, I. Daubechies, and J.-C. Feaveau, "Biorthogonal Bases of Compactly Supported Wavelets," Comm. on Pure and Applied Math., Vol. 45, pp. 485-560, 1992.

[15] Lawrence Rabiner and Biing-Hwang Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1993.