수리 형태론에 의한 새로운 피치 검출기
New Pitch Detectors Using Morphological Filters

蔣 剛 毅*, 崔 太 永**

Gang Yi Jiang*, Tae Young Choi**

ABSTRACT

In this paper, two new morphological pitch detectors, one in time-domain and the other in frequency-domain, are presented. The roles of structuring element in morphological filters are experimentally examined. As a result, it is shown that the new pitch detectors using suitable structuring elements are quite simple, computationally very efficient, and robust than the conventional pitch detectors.

KEYWORDS : Pitch detector, Morphological filter, Speech signal

요 약

본 논문에서는 시간 및 주파수 영역에서 두 가지 새로운 형태론적 피치 검출기를 제안하였다. 형태론적 여파기의 형태소의 역할을 실험적으로 조사하였다. 결론적으로 적절한 형태소를 사용한다면 제안한 방법들은 기존 방법에 비하여 간단하고 속도 면에서도 유리하고 성능 면에서도 강함을 알 수 있었다.

I. INTRODUCTION

A pitch detector is an essential component in a variety of speech signal processing systems, such as speech encoding, speaker identification, speech recognition and speech synthesis [1]. Because of the importance of pitch detection, a wide variety of algorithms on pitch detection of speech signal have been proposed, which include the center clipping autocorrelation method (AUTOC) [2], the cepstral method (CEP) [3], the simplified inverse filtering method (SIFT) [4], the data reduction method (DARD) [5], the parallel processing in time-domain method (PPROC) [6], the spectral flattening linear predictive coding (LPC) method [7], the average magnitude difference function method (AMDF) [8], and so on. Some comments on these methods have been given in [1, 9].

Morphological signal processing is a new research area which is receiving increasing attention. Mathematical morphology which is based on the set-theory provides an approach to the development of nonlinear signal processing operators that incorporate shape information of a signal [10]. The shape of a signal is determined by the values that the signal takes on, and the shape information can be extracted by choosing a suitable structure element to operate on the signal. Morphological filters have successfully been used in image processing and known for robust performance in preserving the shape of signal while suppressing noise [11]. It can also be used to process one-dimensional signal well, an efficient approach to QRS complex detection has been proposed by employing morphological filtering [12].

In this paper, mathematical morphology is used for speech signal analysis. It is computationally very efficient because it uses only morphological operations which are implemented as min/max comparisons. Two new pitch detectors are presented, and the pitches and the spectrum envelopes of speech signals are ob-

*School of Electrical Eng., Ajou University, Korea(Dept. of Electronics Eng., Hangzhou University, P. R. China)
아주대학교 전기전자공학부(중국 항주대학교 전자공학과)
**School of Electrical Eng., Ajou University, Korea
아주대학교 전기전자공학부
접수일자 : 1995년 10월 11일

tained. Experimental results show that they are very efficient and robust.

II. THEORY AND METHOD

2.1 Morphological Filters

Mathematical morphology was introduced by Matheron and Serra, which is set-theoretical methodology for signal processing. It can rigorously quantify many aspects of geometrical structure of signal. The signal transformations of mathematical morphology, which is called morphological filters, are nonlinear operators that locally modify the geometrical features of signal. Here, we are only concerned with the morphological filters with line structuring elements, because they preserve the shapes of signal, allow fast implementation and are robust. Moreover, the line structuring elements are chosen to be symmetric on the origin. If s(n) denotes a one-dimensional signal, where $n \in D = \{0, 1, 2, \ldots, N\}$, and B is a line structuring element of size M, then morphological erosion, $\varepsilon(\ )$, and morphological dilation, $\delta(\ )$, are given by

$$\text{Erosion}\quad \varepsilon(s(n)) = \min\{s(n+m),\ m \in B \text{ and } n+m \in D\} \tag{1}$$

$$\text{Dilation}\quad \delta(s(n)) = \max\{s(n-m),\ m \in B \text{ and } n-m \in D\} \tag{2}$$

Morphological dilation and erosion are two basic morphological operators. Based on them, the opening, $\gamma(\ )$, and the closing, $\xi(\ )$, are defined as

$$\text{Opening}\quad \gamma(s(n)) = \delta(\varepsilon(s(n))) \tag{3}$$

$$\text{Closing}\quad \xi(s(n)) = \varepsilon(\delta(s(n))) \tag{4}$$

Further, the open_close, $\varphi(\ )$, and the close_open, $\psi(\ )$, are defined as

$$\text{Open\_close}\quad \varphi(s(n)) = \xi(\gamma(s(n))) \tag{5}$$

$$\text{Close\_open}\quad \psi(s(n)) = \gamma(\xi(s(n))) \tag{6}$$

The opening (resp. closing) is an important morphological filter which simplifies signal by removing the peak (resp. valley) components which do not fit within the structuring element. If the simplification has to deal with both peak and valley components, the open_close and the close-open operators can be used.

2.2 Pitch Detector in Time-Domain Using Morphological Filter

The conventional AUTOC method using the center clipping autocorrelation can not eliminate sufficiently the formant components in transition region because a fixed threshold in each frame is used to reduce formant components. To overcome the problem of fixed threshold and to obtain adaptive threshold, we propose a new pitch detector using a morphological filtering technique in time-domain as shown in Fig. 1. In the figure, morphologically filtered speech signal, f(n) and g(n), play the role of the adaptive thresholds able to clip the formant components in transition regions. And the unvoiced/voiced decision outputs a switch sign, u/v, to control the states of clipper, according to f(n) and g(n). If it is a voice signal v, s(n) will be positively and negatively clipped by the adaptive thresholds. Autocorrelation of clipped signal is fast calculated and the pitch period of s(n) is obtained exactly. If s(n) is unvoiced signal or silent and it is not necessary to detect pitch. To extract the adaptive positive and negative thresholds of s(n), we use open-close and close-open filters as follows

$$f(n) = \varphi(s(n)) \tag{7}$$

$$g(n) = \psi(s(n)) \tag{8}$$

Fig. 2 shows examples of voice signal filtered by morphological filters. In the figure, (a) is s(n), (b) is center clipped s(n) by the adaptive thresholds of (c) and (e), (c) is the adaptive negative thresholds obtained by $\varphi(s(n))$, (d) is clipped s(n) by the adaptive negative thresholds in (c), (e) is the adaptive positive thresholds obtained by $\psi(s(n))$, and (f) is clipped s(n) by the adaptive positive thresholds in (e).

A comparison between the new clipping and the conventional center clipping is shown as Fig. 3. (a) is s(n), (b) is clipped s(n) with the conventional center clipping, (c) is clipped s(n) with the new clipping (M=91), (d) is clipped s(n) with the new clipping (M=21), and (e) is clipped s(n) with the new clipping (M=201). It is obvious that the new clipping is better than the conventional center clipping. Therefore, the new pitch detector is better than the conventional pitch detector using the center clipping autocorrelation. However, it is important how to determine a suitable structuring element of morphological filter in the new clipping in order to get robust

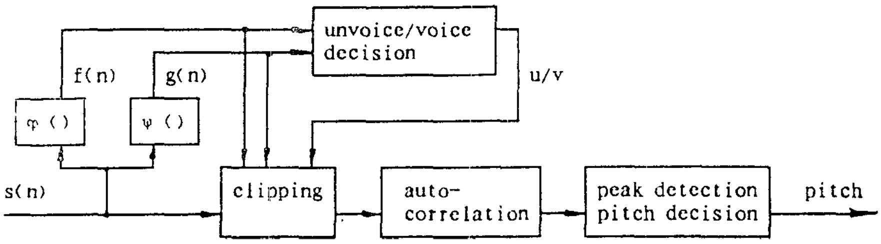


Fig 1. Pitch detector in time-domain based on morphological filters

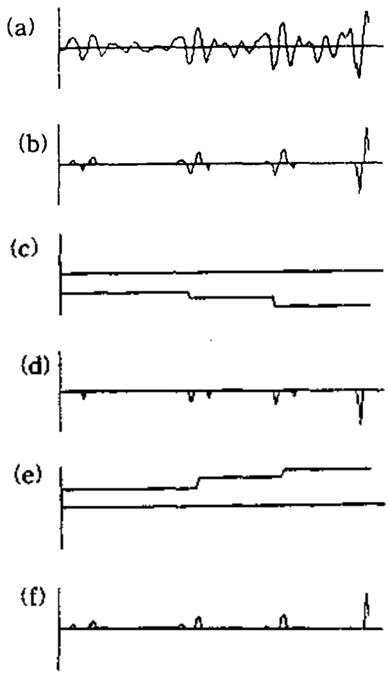


Fig 2. Filtered voice signal waveforms by morphological filters
(a) s(n),
(b) center clipped s(n) by the adaptive thresholds of (c) and (e),
(c) the adaptive negative thresholds obtained by $\varphi(s(n))$,
(d) clipped s(n) by the adaptive negative thresholds in (c),
(e) the adaptive positive thresholds obtained by $\psi(s(n))$,
(f) clipped s(n) by the adaptive positive thresholds in (e).

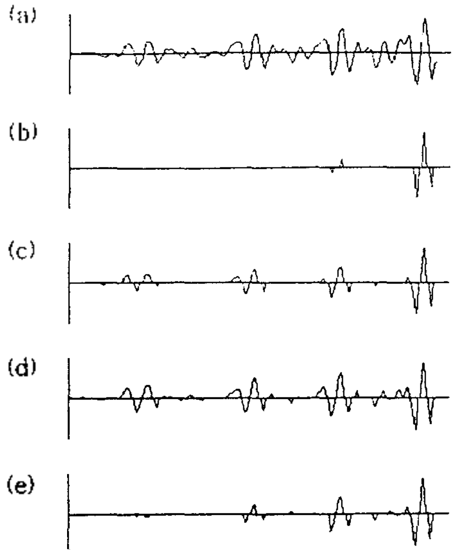


Fig 3. Comparison of clipped waveforms
(a) original speech signal s(n),
(b) conventional center clipped waveform,
(c) morphologically clipped waveform with M = 91,
(d) morphologically clipped waveform with M = 21
(e) morphologically clipped waveform with M = 201

peak information. In Fig. 3(d), M is too small to eliminate well formant information in transition regions. On the contrast, M is too large in Fig. 3(e). M can be selected about 1.5 times pitch period of speech signal.

2.3 Pitch Detector in Frequency-Domain Using Morphological Filter

Fig. 4 shows a new pitch detector in frequency-domain using morphological filter. Speech signal s(n) is inputted and pre-processed with a window such as Hanning window, and its Fourier magnitude spectrum is computed and used to determine which s(n) is silent, unvoiced or voiced. If the output of the silent/unvoiced/voiced (s/u/v) is s, it implies that s(n) is silent and it is not necessary to detect speech features. If the s/u/v is u, it implies that s(n) is unvoiced signal and has only formant information; if the s/u/v is v, it implies that s(n) is voiced signal with formant information and pitch to be extracted. Then, the log magnitude spectrum of s(n), x(k), is computed and filtered morphologically to find an estimate of log magnitude spectral envelope f(k) as noted in Eq. (9). Since g(k) in Eq. (10) is the residue of filtered log magnitude spectral, g(k) can be

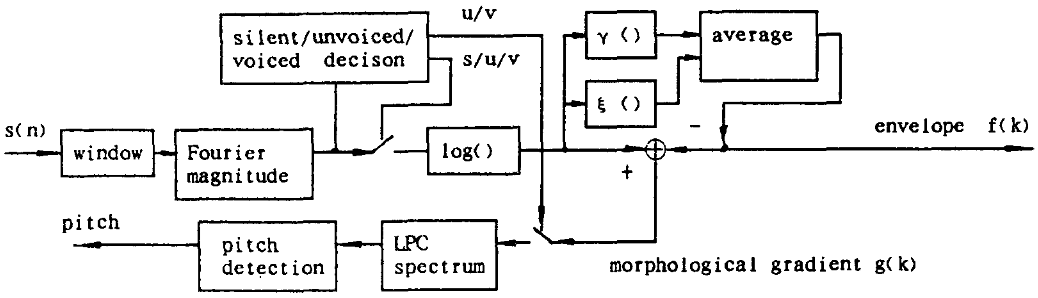


Fig 4. Pitch detector in frequency-domain using morphological filters

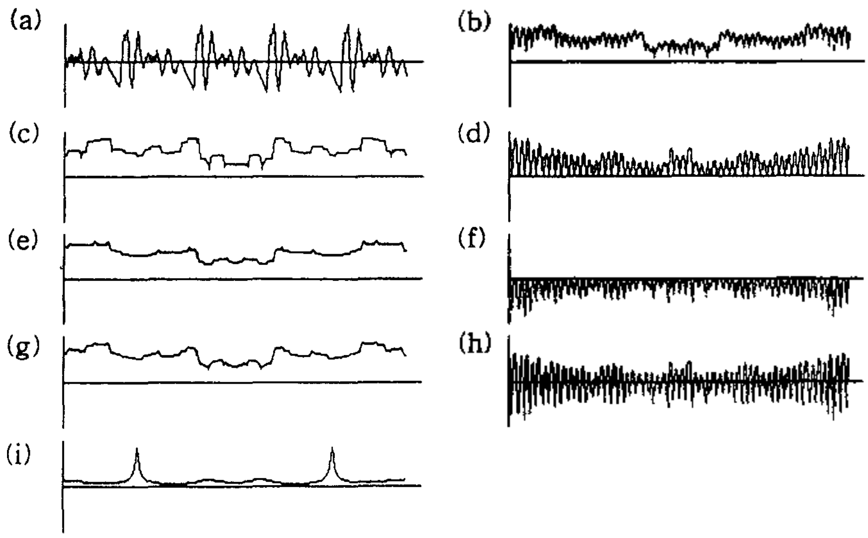


Fig 5. Morphological filterings in frequency-domain for pitch detection
(a) s(n), (b) log magnitude spectrum x(k),
(c) $\gamma(x(k))$, (d) $x(k)-\gamma(x(k))$,
(e) $\xi(x(k))$, (f) $x(k)-\xi(x(k))$,
(g) f(k), (h) g(k),
(i) LPC spectrum of g(k).

used to estimate the pitches of s(n), and the peak of its LPC spectrum corresponds to the pitch period if the parameter of LPC are reasonably selected.

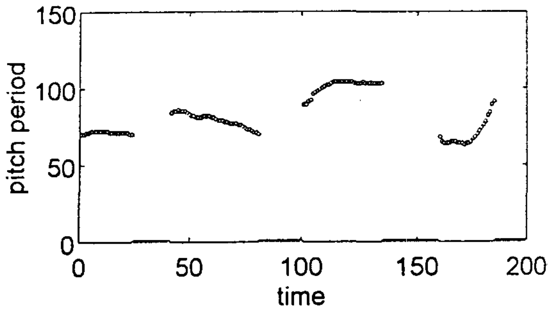
$$f(k) = 0.5 \times \{\gamma(x(k)) + \xi(x(k))\} \quad (9)$$

$$g(k) = x(k) - f(k) \quad (10)$$

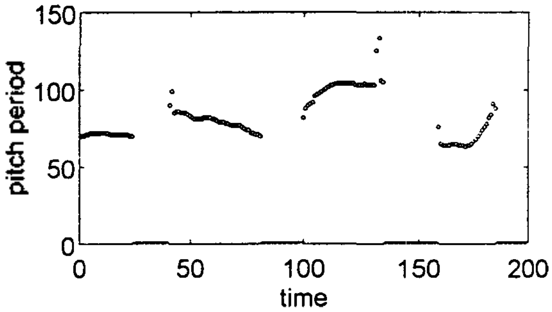
Fig. 5 illustrate pitch detecting technique using morphological filters in frequency-domain, (a) is s(n), (b) is the log magnitude spectrum x(k), (c) is $\gamma(x(k))$, (d) is $x(k)-\gamma(x(k))$, (e) is $\xi(x(k))$, (f) $x(k)-\xi(x(k))$, (g) envelope of x(k), f(k), (h) residue g(k), and (i) is LPC spectrum of g(k), where the location of the peak is the pitch period of s(n). It is similarly important how to determine a structuring element for a good estimate of spectrum envelope.

Ⅲ. EXPERIMENTS AND ANALYSES

Some results of the new pitch detector in time-domain have already been shown in the Fig. 2 and Fig. 3. Fig. 6 shows pitch contours for a long segment of speech signal with the new pitch detector in time-domain and the conventional AUTOC method. It is evident that the new pitch detector in time-domain is better than the conventional method.



(a) pitch contours of the new pitch detector in time-domain



(b) pitch contours of the AUTOC method

Fig 6. Pitch contours of speech signal extracted by the new pitch detector in time-domain and the AUTOC method

Fig. 7 shows an analysis of detecting features of speech signal in frequency-domain. (a) is s(n), (b) is the log magnitude spectrum x(k), (c) is envelope of x(k) with M=5, f(k), (d) is envelope of x(k) with M=45, f′(k), (e) is residue g(k), g(k)=x(k)−f(k), (f) is residue g′(k), g′(k)=x(k)−f(k), (g) is the LPC spectrum of g(k) with P=8, (h) is the LPC spectrum of g′(k) with P=8, (i) is the LPC spectrum of g(k) with P=12, and (j) is the LPC spectrum of g′(k) with P=12, where P is the order of LPC.

In Fig. 7(c), the size of structuring element, M, is too large to extract the spectrum envelope perfectly, and some spectrum envelope information is still remained in g(k). Similarly, the order of LPC, P, is so large that the LPC spectrum of g(k) gets bad in Fig. 7(i). For this example of Fig. 7, the size M can be determined to be 5 or 7, while the order of LPC, P, is chosen to be from 6 to 8.

Ⅳ. CONCLUSIONS

The morphological filters are used to simplify speech signal and its spectral data, Two new pitch detectors

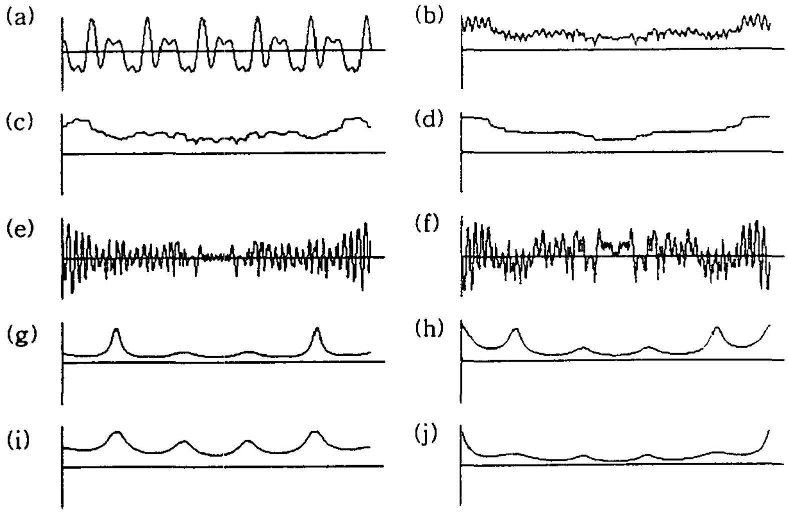


Fig 7. Analysis of new pitch detection in frequency-domain
(a) s(n), (b) log magnitude spectrum x(k),
(c) f(k) with M=5, (d) f′(k) with M=45,
(e) residue g(k)=x(k)−f(k),
(f) residue g′(k)=x(k)−f′(k),
(g) LPC spectrum of g(k) with P=8,
(h) LPC spectrum of g′(k) with P=8,
(i) LPC spectrum of g(k) with P=12,
(j) LPC spectrum of g′(k) with P=12.

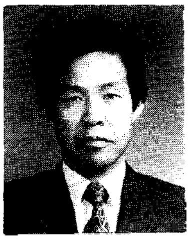
are realized, which are intuitive, computationally efficient, and apt to parallel implementation. Some experiments are carried out to compare the performance of the new methods with the conventional ones, and to analyze the effect of structuring elements in the new methods. Experimental results show that the new pitch detectors using suitable parameters are more powerful and robust than the conventional methods.

REFERENCES

1. S. Furui, *Digital Speech Processing, Synthesis, and Recognition*, Marcel Dekker, Inc. 1989.
2. L.R. Rabiner, "On The Use of Autocorrelation Analysis for Pitch Detection," IEEE Trans. ASSP-25, pp. 24-33, 1977.
3. A.M. Noll, "Cepstrum Pitch Determination," J. Acoust. Soc. Am., 41, pp. 293-309, 1967.
4. J.D. Markel, "The SIFT Algorithm for Fundamental Frequency Estimation," IEEE Trans. AU-20, pp. 367-377, 1972.
5. B. Gold, et al., "Parallel Processing Techniques for Estimation Pitch Periods of Speech b in the Time Domain," J. Acoust. Soc. Am., 46, pp. 442-448, 1969.
6. N.J. Miller, "Pitch Detection by Data Reduction," IEEE Trans. ASSP-23, pp. 72-79, 1975.
7. M.M. Sondhi, "New Methods of Pitch Extraction," IEEE Trans. AU-16, pp. 262-266, 1968.
8. M.J. Ross, et al., "Average Magnitude Difference Function Pitch Extractor," IEEE Trans. ASSP-22, pp. 353-362, 1974.
9. R.W. Schafer, J.D. Markel, *Speech Analysis*, IEEE Press, 1979.
10. O.R. Giardinn, E.R. Dougherty, *Morphological Methods in Image and Signal Processing*, Prentice-Hall, Inc. 1988.
11. P. Maragos et al., "Morphological Systems for Multidimensional Signal Processing," Proceedings of the IEEE, 78, pp. 690-709, 1990.
12. P.E. Trahanias, "An Approach to QRS Complex Detection Using Mathematical Morphology," IEEE Trans. BME-40, pp. 201-205, 1993.

Gang Yi Jiang was born in Zhejiang, P. R. China, in 1964. He received the B.S. and M.S. degrees in Electronics Engineering from Zhejiang Radio and Television University and Hangzhou University in 1985 and 1992, respectively.

Since 1992, he has been assistant Professor of Hangzhou University. He is currently working with Professor Tae Young Choi as a researcher with a contract of one year in Ajou University. His general interests was in the area of digital speech signal processing and he is currently doing research on use of mathematical morphology for automatic vision system such as traffic sign recognitions.



Tae Young Choi was born in Kyungnam-do, Korea, in 1950. He received the B.S. and M.S. degrees in Electronics Engineering from the Seoul National University, Seoul, Korea, in 1974 and 1978, respectively, and D.E.A. and Docteur-Ingênieur degrees from University of Aix-Marseille III, Marseille, France, in 1980 and 1982, respectively.

Since 1983, he has been with Ajou University, where he is currently a professor of School of Electrical Engineering and director of Academy of University-Industry Cooperation. His general intersests are in the area of digital signal processing and he is currently doing research on mathematical morphology and its applications.