2000년도 한국음향학회 학술발표대회 논문집 제 19 권 제 2(s) 호

Dyadic Wavelet Transform 방식의 Pitch 주기결정

김 남훈 윤기범 고 한석

고려대학교 전자공학과
서울시 성북구 안암동 5가 1번지

A Stable Pitch Determination via Dyadic Wavelet Transform (DyWT)

Namhoon Kim Gibum Yoon Hanseok Ko

Dept. of Electronics Engineering, Korea University
5ka-1 Anam-Dong Sungbuk-Ku, Seoul 136-701, Korea
Tel: +82-2-926-2909, Fax: +82-2-3291-2450
E-mail:nhkim@ispl.korea.ac.kr, hsko@korea.ac.kr

Abstract

This paper presents a time-based Pitch Determination Algorithm (PDA) for reliable estimation of Pitch Period (PP) in speech signal. In proposed method, we use the Dyadic Wavelet Transform (DyWT), which detects the presence of Glottal Closure Instants (GCI) and uses the information to determine the pitch period. And, the proposed method also uses the periodicity property of DyWT to detect unsteady GCI. To evaluate the performance of the proposed methods, that of other PDAs based on DyWT are compared with what this paper proposed. The effectiveness of the proposed method is tested with real speech signals containing a transition between voiced and the unvoiced interval where the energy of voiced signal is unsteady. The result shows that the proposed method provides a good performance in estimating the both the unsteady GCI positions as well as the steady parts.

1. Introduction

Accurate PP detection can improve the performance of many speech applications requiring signal processing, e.g speech recognition, speaker verification, and lip-synch since PP provides a discerning feature useful for classifying phonemes. In short, it can be used as an indicator for phoneme segmentation. However, reliable and accurate PP detection poses a formidable challenge because of the following reasons. First of all, the human vocal tract is very flexible and as a result its characteristics vary widely depending on the individual. Second, even though the information of PP comes from the same speaker, it can be changed by the emotional state of the speaker. Finally, PP can be affected by the individuals speaking style, e.g. accent or intonation.

This paper mainly focus on the event-based PDA, more specifically, the one based on DyWT. The wavelet transform is a multi-scale analysis which has been shown to be very well suited for speech signal processing in that it decomposes speech signal into a series of band-pass components. It is similar to the way how human ears process sound. Kadambe and Bordreaux-Bartels used the assumption that as the GCI occurs in speech signal, maximums also happen in the adjacent scales of the wavelet transform. Therefore, the classification between the unvoiced and the voiced can not be done by only comparing the maximum amplitude of the DyWT with some threshold, but also checking whether the local maxima of the DyWT is identical across the two scales [1, 4]. However, although Kadambe's work shows relatively good performance for detecting the steady GCI, unfortunately, it does not provide reliable results for estimating the unsteady GCI positions such as the transition between voiced and the unvoiced or the beginning and ending of the voiced where the energy of it is not steady. The Wendt's method [2], another PDA based on DyWT, used a filtering function, which is linked to the bandwidth properties of the wavelet transform at different scales. it shows the better performance in detecting unsteady GCIs in comparison to the Kadambe's method, because it does not use a threshold to detect GCIs. But it also suffers from the unwanted maxima, which incur PP detection errors. As a remedy to the problem discussed in this paper, we suggest new PDA, which explores the compensation of the above-mentioned two PDAs based on DyWT to be free from PP estimation errors.

This paper is organized as follows. A brief description of the adopted DyWT is presented in Section 2, Section 3 discusses detail mention of the problem of conventional PDAs based on DyWT, Section 4 gives the proposed PDA. Then, Sections 5 and 6 presents the representative results and conclusions respectively.

2. Background

The DyWT of a signal $x(t)$ is defined as, [1, 3]

$$DyWT_x(b,2^j) = \frac{1}{2^j}\int_{-\infty}^{\infty} x(t)g^*\left(\frac{t-b}{2^j}\right)dt = x(t) * g^*_{2^j}(t) \quad (1)$$

where $g_{2^j}(t)$ is the dilated and scaled version of wavelet. Note that b and 2^j are limited to integer, and that $DyWT(b,2^j)$ is the Dyadic wavelet transform coefficients representing the wavelet transform at each 2^j scale.

The DyWT acts as a constant-Q filter bank and splits the signal into band-pass components. This is very useful in the analysis of characteristics of the signal, which are localized in frequency. In realization of Kadambe's method in this paper, cubic spline wavelet is used, and $DyWT(b,2^j)$ at scale $j = 3, 4$ are used to estimate the GCI position. Mallat's algorithm is also utilized for a fast implementation of DyWT [2]. On the other hand, haar wavelet function at scale $j = 3$, and scaling function at scale $j = 6$ are used to implement Wendt's PDA [2]. Finally, in the proposed method, both wavelet function and scaling function at scale $j = 5$ are adopted to construct filtering function as a similar fashion in the Wendt's.

3. The Conventional PDA based on DyWT

3.1 S. Kadambe's Method

Kadambe's method [1, 4], which compares the maximum amplitude of the DyWT with a certain threshold level T in addition to checking whether the local maxima of the DyWT correlates across two scales. However, we have found through extensive experimentation that the fixed threshold is too strong in some low energy voiced segment as in the transition between voiced and the unvoiced or the beginning and ending of the voiced where the energy of it is not steady. As shown in Fig. 1, we failed the accurate estimation of GCI position with their method.

3.2 C. Wendt's Method

In Wendt's method, the idea is to use a wavelet with the derivative properties, described by Mallat [3], that also combines the bandwidth properties of the wavelet transform at different scales. Since the frequency range of voiced speech is between *30 - 500Hz*, a filtering function is constructed to cover the similar range of it by using both low-pass scaling function and high-pass wavelet function. The filtering function $\rho(t)$ is obtained as below,

$$\rho(t) = \psi_{K_a}(t) * \phi_{K_b}(t)$$
$$s'(t) = s(t) * \rho(t) \quad (2)$$

where $\psi_{K_a}(t)$ and $\phi_{K_b}(t)$ are wavelet and scaling function, K_a and K_b are corresponding upper- and lower- bound scale, $s'(t)$, $s(t)$, and $\rho(t)$ are the filtered signal, speech signal and derivative filtering function, respectively [2].

However, as shown in Fig. 2, even though speech signal is filtered out in the range of about *30 - 500Hz*, the unwanted local maxima still remain to cause GCI detection error. Furthermore, although it is possible to get rid of unwanted local maxima as increasing scaling index, in this case, there exists big difference between GCIs obtained and real GCIs.

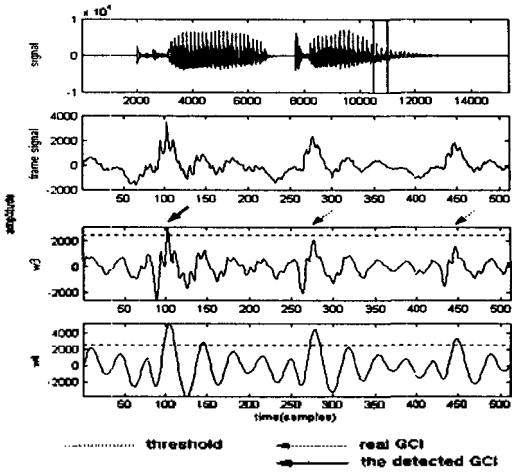

Fig.1 : Detection of GCI's with Kadambe and Bordreaux-Bartels's method

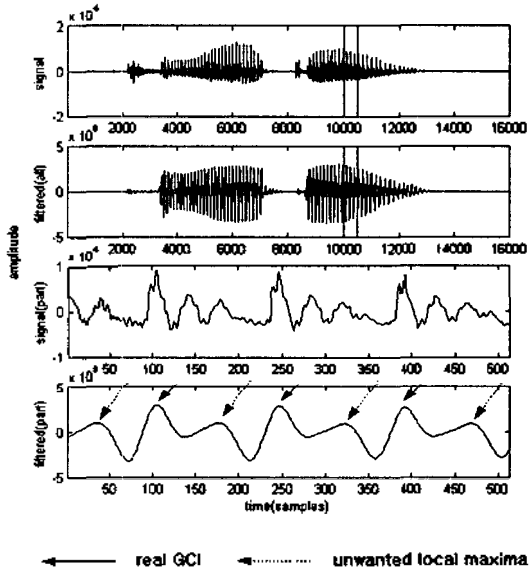

Fig.2 : Detection of GCI's with Wendt and A. Petropulu's method

4. The Proposed Method

The proposed method also follows the C. Wendt and A. Petropulu's method [2], which construct filtering function with both low-pass scaling function and high-pass wavelet function to cover the range of the voiced, that is, about *30 - 500Hz*. But, unlike the Wendt's method, We utilize filtering function, which is made of cubic spline wavelet at scale $j = 5$, and also cover the frequency range between *30 - 500Hz* approximately. The proposed method we present in the following subsections is consisted of two important stages; variant threshold stage and voiced/unvoiced detection stage.

4.1 Variant threshold stage

The threshold is imposed according to the detected local maxima. As shown in the Fig. 3, once the local maxima have been detected, the proposed PDA takes its threshold from the lowest local maxima value to the highest. And, simultaneously, thresholded local maxima are compared with this variant threshold. The adopted threshold is given as below,

$$Threshold = s'(n) \qquad 1 \le n \le M \qquad (3)$$

where $s'(n)$ is the frame signal and M is the number of local maxima within the frame.

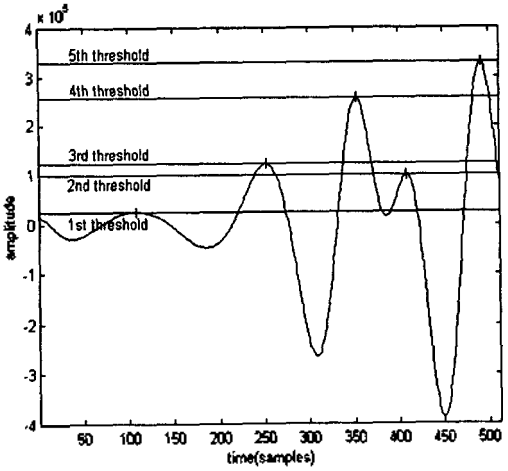

Fig.3 : Variant threshold

4.2 Voiced/unvoiced detection stage

With given thresholded local maxima, it is affirmed that they are real GCI or not. In the case of steady GCI position, we evaluate each PP and mean of it for all the thresholded local maxima. If the analyzed frame is voiced segment, each PP get a value, which is almost same to mean of PP within relatively small error. However, when it comes to the unsteady GCI position, it needs another treatment. As noted in (4), as scale k increases, the bandwidth of filtering function, which is composed of both $\psi_K(t)$ and $\phi_K(t)$, decreases as much as 2^k in frequency domain. In other words, the signal is filtered at the higher scale, the periodicity information is higher in that quasi-periodic factors in signal get filtered out, on the contrary, the gap between real GCI and candidates grows up.

$$\begin{aligned} s'(t) &= s(t) * \rho(t) \\ &= s(t) * \psi_k(t) * \phi_k(t) \\ S'(w) &= 2^{-k} S(w) \cdot \Psi(2^{-k} w) \cdot \Phi(2^{-k} w) \end{aligned} \qquad (4)$$

where $s'(t)$, $s(t)$, $\rho(t)$, $\psi_K(t)$, and $\phi_K(t)$ are filtered signal, analyzed signal, filtering function, wavelet, and scaling function, respectively. And $S'(w)$, $S(w)$, $\Psi(2^{-k} w)$, and $\Phi(2^{-k} w)$ are Fourier transform of $s'(t)$, $s(t)$, $\psi_K(t)$, and $\phi_K(t)$, respectively.

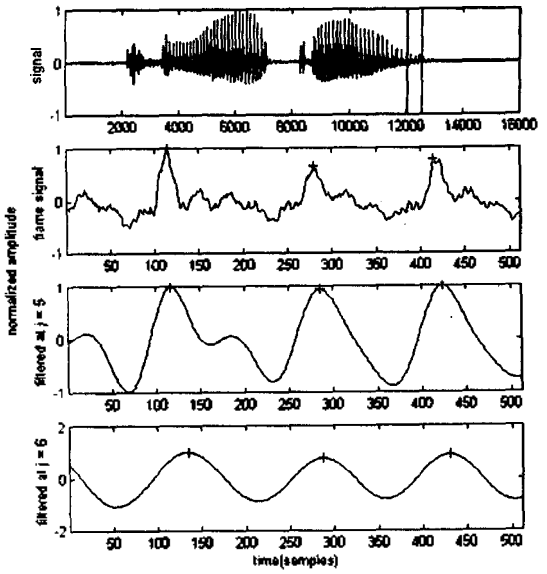

Fig.4 : Comparison of GCI position in the filtered signal at scale j = 5 and 6

Accordingly, We use the periodicity information to confirm whether it is real GCI or not at higher scale filtered signal, that is, filtered signal at scale $j = 6$. But, in practice, we do not separate steady GCI and unsteady GCI, but test all GCIs with the above-mentioned method. The comparison of GCI position in the filtered signal at scale $j = 5$ and 6 is depicted in Fig. 4 and The overall procedure of the proposed method is depicted in Fig. 5

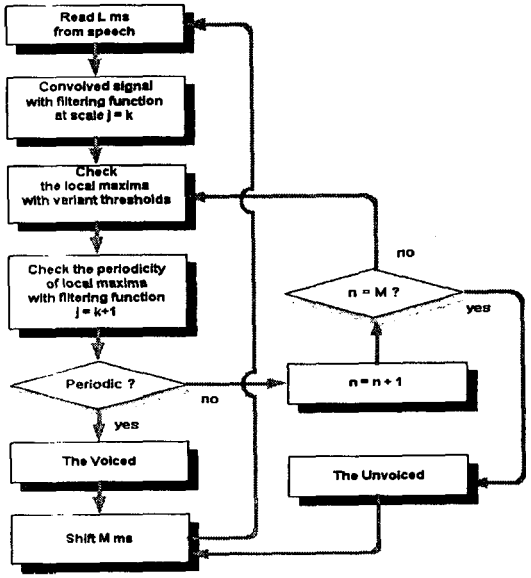

Fig.5 : Flowchart of the proposed method

5. Results and Discussion

We provide a performance comparison of the three PDAs on real speech signals; the proposed method, Kadambe's method, and the Wendt's method. The speech database used is composed of four sentences, also uttered by four male speakers, and sampled at 16kHz. In all the PDAs, We use the following

measurements; GGDE(Gross GCI Determination Error, say, errors greater than 1ms), FGDE(Fine GCI Determination Error, say, errors below than 1ms), and VDE(Voicing Determination Error, over 30ms). Table 1 shows the overall results.

	GGDE(%)	FGDE(%)	VDE(%)
the Kadambe's	9.16	3.31	3.69
the Wendt's	0.94	55.80	0.65
the Proposed	0.91	2.20	0.65

Table. 1: GCI detection results

The table.1 and Fig. 6 tells that the proposed method shows the best performance in comparison of other two. It is also noted that in the case of the Kadambe's method, it shows the biggest GGDE due to the fixed threshold, and the Wendt's method also shows the biggest FGDE due to unwanted local maxima.

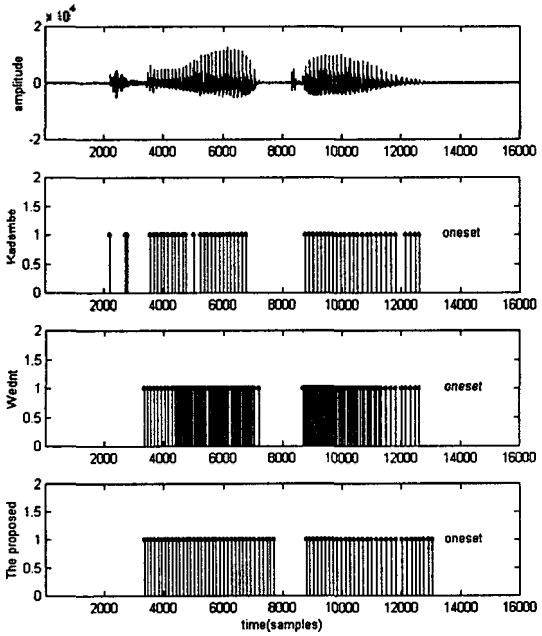

Fig.6 : Speech signal, "Chung-Wa-Dae" and its GCI detection results

6. Conclusions

We presented an effective PDA, which is based on the DyWT. By comparing with other prominent methods based on DyWT. We have shown that the proposed method is better in terms of performance especially in those speech segments where a transition between voiced and unvoiced take place, and where the beginning or ending of the voiced exists in the analyzed speech segment.

Acknowledgment : *this work has been supported by the University Basic Research grants from the Ministry of Information & Communication. The authors gratefully acknowledge the Ministry's support.*

References

[1] Shubha 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.

[2] Christher Wendt and Athina P. Petropulu, *"Pitch Determination and Speech Segmentation Using the Discrete Wavelet Transform,"* Proc. IEEE International Symposium on Circuit and System, 1996, vol. 2, pp. 45-48

[3] Stephane Mallat and Sifen Zhong, *"Characterization of Signals from Multiscale Edges,"* IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.14, No.7, pp.710-732, July 1992.

[4] Shubha Kadambe and G.F. Boudreaux-Bartels, *"A Comparison of a Wavelet Functions for Pitch Detection of Speech Signals, "* IEEE, pp.449-452, 1991.