Improvement of Pitch Detection using Signal Specific Analyzing Wavelet

Yoshifumi Chisaki, Tsuyoshi Usagawa and Masanao Ebata

Dept. of Computer Science, Faculty of Engineering, Kumamoto University, Kumamoto, Japan

Abstract: This study deals with robust pitch detection in noisy environments. Harmonic wavelet transform using signal specific analyzing wavelet is used to obtain a pitch where noise exists. As the result, a ratio of pitch detection is improved by the proposed method compared with a general wavelet transform using Gabor function. The maximum difference between two methods is 48% at SNR of -5dB.

INTRODUCTION

An algorithm for robust and precise pitch detection is expected for many signal enhancement systems in composite sound source condition. In those studies, increasing error of estimated pitch against true pitch grows down a total performance of system so much. In our previous study, one of speech enhancement systems in multiple speaker condition was proposed [1], and that system had the same characteristics. The system was composed of three blocks; spectrum analyzer, pitch trackers, and adaptive filter banks. Since the system deals with speech, pitch estimation block makes use of information of both frequency and power of harmonic components. And the information of harmonics was given by spectrum analyzer based on Fast Fourier Transform (FFT). Although FFT is very useful to analyze a signal, time and frequency resolution is not sufficient to improve the quality of enhanced speech. Wavelet transform is one of useful transforms to obtain a precise pitch because the frequency resolution in lower range is better than that of FFT. And the restriction for the analyzing wavelet is only the admissibility condition, so the characteristics of wavelet transform can be changed with ease. This flexibility and efficient analysis for speech is focused on for substitution of FFT as spectrum analyzer. Let the analyzing wavelet assume to have characteristics of target speech, peak of power at a pitch frequency is acumulate in spectrogram because power of harmonic components is summed up. This analyzing wavelet which has characteristics of a target signal is called signal specific analyzing wavelet hereafter. Although this idea gives robustness for pitch detection, information of both amplitude and phase for harmonic components appears as ripple of power along time axis. This phenomenon makes pitch tracking difficult when this method is applied to the system of the previous study. For solving this problem and obtaining more precise pitch, modified wavelet transform using signal specific analyzing wavelet is proposed as harmonic wavelet transform.

THEORY

In this section, harmonic wavelet transform is derived. An ordinary continuous wavelet transform is defined as,

\[ W(b,a) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{a}} \psi \left( \frac{x-b}{a} \right) f(x) dx, \]  

where the parameter \( b \) and 1/a are corresponding to time and frequency, respectively. Also \( f(x) \) is analysis signal, and \( \psi(x) \) is analyzing wavelet. Generally, a function which is derived from mathematical approach, such as Gabor, is used as the analyzing wavelet. And the analyzing wavelet performs as a band pass filter. Thus power of a target signal at one frequency is interfered by powers of other signals at the same frequency, so acumulate peak of power at pitch frequency cannot be observed. Signal specific analyzing wavelet which has a characteristics of a target signal is used to obtain acumulate peak of power. Signal specific analyzing wavelet is constructed by following two steps; a continuous signal is made by using a series of a fundamental period of target signal, and the continuous signal is weighted by hamming window function. Length of window function depends on the uncertainty principle. Equation (1) is modified to adopt this specific signal to analyzing wavelet as following.

Let us assume a function \( g(a,b,x) \) as a part of Eq.(1) as the following:

\[ g(a,b,x) = \frac{1}{\sqrt{a}} \psi \left( \frac{x-b}{a} \right). \]  

Thus Eq.(1) can be represented as the following;

\[ W(b,a) = \int_{-\infty}^{\infty} g(a,b,x) f(x) dx. \]  

The parameter \( b \) and 1/a mean time and frequency, respectively, so Eq.(1) can be expressed with replacing both
parameters to time parameter $t$ and frequency parameter $f$,

$$W(t,f) = \int g(f,t,x)f(x)dx.$$  \hspace{1cm} (3)

The signal specific analyzing wavelet is applied to $g(f,t,x)$. Spline function is used to vary analysis frequency $f$. Although wavelet transform can be performed by Eq. (3), a ripple of power along the time axis occurs in case that a general wavelet transform is performed when the signal specific analyzing wavelet is used as analyzing wavelet. The reason of this phenomenon is that the analyzing wavelet has information of both amplitude and phase for harmonic components [2]. This phenomenon makes pitch tracking difficult when the power is taken into account to track a pitch. For solving this problem, harmonic wavelet transform is defined as following:

$$w_s(t,f) = \frac{1}{2\pi} \int g(t+T,f,x)f(x)dx$$  \hspace{1cm} (4)

where $T$ is a period of analysis frequency.

**EXPERIMENTS**

Experiments are performed to examine whether harmonic wavelet transform is more robust against noise than a general wavelet transform. This means whether the proposed method is adequate for analyzer module for our previous speech enhancement system or not. Thus the undermentioned thresholds are more strict than that of FFT. In this experiment, ratio of pitch detection is used as index of robustness. The experiment of pitch detection is performed in concurrent situation. Vowels (/a/, /i/, /u/, /e/, /o/) uttered by 10 females and 10 males are used as a target signal, respectively. And the other signal is a band limited noise where the range of frequency is set to overlap around the pitch frequency of a target signal. The SNR is calculated by power of overlapped frequency band. Sampling frequency is 10kHz, and an evaluated duration of utterance is 100ms. A pitch obtained by Gabor at SNR of 0 dB is used as a criterion to judge whether "detected" or not. A parameter $\sigma$ which corresponds to quality factor of filter is set to 4. The "detected" is defined so that the absolute frequency difference between a criterion pitch and the frequency located where the peak of power exists is less than 3Hz. This threshold is less than a resolution of frequency for spectrum analyzer based on FFT. Figure 1 shows results of pitch detection ratio obtained by each method. And the ratio is an average of 10 females and 10 males for each vowel. The left bars show a ratio of pitch detection obtained by Gabor function and right ones are that obtained by harmonic wavelet transform. The ratio obtained by harmonic wavelet transform is higher than one obtained by Gabor function in all vowels. In case of female /a/, a maximum difference of ratio 36.2% is obtained. Table 1 shows the results of the maximum and minimum difference in all vowels when SNR is varied from -5dB to 5dB with 5dB step. The maximum difference was 48.0% at SNR of -5dB. However, ratio obtained by Gabor is better than that obtained by the proposed method at SNR of 5dB.

**REFERENCES**


<table>
<thead>
<tr>
<th>SNR</th>
<th>Max(%)</th>
<th>Min(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2.9</td>
<td>-20.9</td>
</tr>
<tr>
<td>0</td>
<td>36.2</td>
<td>13.6</td>
</tr>
<tr>
<td>-5</td>
<td>48.0</td>
<td>11.3</td>
</tr>
</tbody>
</table>

**FIGURE 1.** Ratio of pitch detection at SNR of 0dB.