ウェーブレット変換による音声解析
- 中国北京官話の子音特徴解析への応用 -

川頭　信之

農業生物資源研究所　〒305-8602 茨城県つくば市観音台2-1-2

E-mail: kawagash@affrc.go.jp

あらまし
本研究では音声解析にウェーブレット変換を導入し、その有効性を検証し、新しい音声解析方法を提示するのが目的である。スペクトログラフは窓フーリエ変換を用いているが、ウェーブレット変換はフーリエ変換とは異なり、周波数と位置情報を同時に捉えることが可能である。ウェーブレット変換はスペクトル変化の激しい言語音の特徴抽出に有効であると考え、中国語北京官話の唇音[pa, pʰa]にウェーブレット変換を施し、無気音・有気音の子音特徴を調べた
キーワード　ウェーブレット変換、北京官話、子音、有気音、無気音、唇音、フーリエ変換

Speech Analysis with Wavelet Transform
- Its Application to Consonantal Feature Analysis in Chinese Mandarin -

Nobuyuki Kawagashira

National Institute of Agrobiological Sciences, Kannondai 2-1-2, Tsukuba, Ibaraki, 305-8602 Japan

E-mail: kawagash@affrc.go.jp

Abstract　In this study, I applied wavelet transform to speech analysis and showed its effectiveness to speech analysis. Wavelet transforms can detect frequency and positional information at the same time thou Fourier Transforms cannot. Wavelet transforms are useful for detecting linguistic sound features which have their rapid spectral change in time. Wavelet transforms are applied to consonantal feature analysis of Chinese Mandarin aspirated/unaspirated labials [pa, pʰa]. I will try to show that wavelet transforms are more useful than Fourier-based spectrograph for speech analysis.
Keyword　wavelet transform, Chinese Mandarin, consonant, aspirated, unaspirated, labial, Fourier

PAPER

Speech Analysis with Wavelet Transform: Its Application to Consonantal Analysis in Mandarin Chinese

Nobuyuki KAWAGASHIRA

1. Introduction

1.1 Spectral Analysis with Fourier Transform

In the signal processing field, Fourier transform is popular for analyzing wave. It is appropriate for detecting frequency element of the total wave form, but not for the wave changing its spectra in time. Windowed Fourier Fourier transform is used for spectrograph by partitioning the wave form into smaller intervals (windows). Then the longer correlations between wave subsections are lost in Fourier transform.

1.2 Usefulness of Wavelet Transform

On the other hand, wavelet transform has several different feature from the traditional time-frequency analysis or Fourier analysis because it expands signals into the time-frequency plane for analysis. Wavelet method offers both frequency informations (spectrum) and locational one at the same time. Wavelet method is now increasing its importance for signal processing. In this study, multiresolution analysis of wavelet transform is used, which can decompose data in the multiple scale (wavelength) analysis. I will introduce a new method for speech analysis using wavelet transform.

1.3 Mandarin Chinese

Chinese language is called the second largest population of speaker next to English in the world. Chinese mandarin is taught as common language in China. It is a tonal language, having 4 main tones such as *mā*, *má*, *mǎ* and *mà*, and one light tone called *qīngshēng*. In this study, the first tone (high flat) is used for avoiding confusion caused by tonal difference. Recognition of aspirated/unaspirated consonant is important for speaking Mandarin Chinese. According to the speech recognition study, consonantal characteristics is determined by the spectral change in time, which is different from the vowel recognition. The phonetical feature of aspirated/unapirated consonant is still unclear until now in spite of previous phonetical researches. [3] So labial plosive sounds *bā* [pā] and *pā* [pʰā] are chosen for showing the usefulness of wavelet transform.

2. Methods

2.1 Wavelet Transform

Wavelet analysis is a method decomposing wave function $f(x)$ into frequency and locational element by a specific small wave called mother wavelet $\psi(x)$. Scaling is a fundamental operation used in the wavelet transform. The analysis function for the wavelet transform is defined as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \quad a \in \Re$$

Wavelet transform is defined using this scaling function $\psi(x)$ as

$$W_{\phi}(a,b) = \int_{-\infty}^{\infty} f(x)\overline{\psi_{a,b}(x)}dx.$$

2.2 Discretization of Wavelet Transform

When we consider plotting the wavelet transform values on 2-dimensional coordinate system, we can choose non-overlapping coordinate by $(b, 1/a)$ on the plane. Replacing the coordinate $(b, 1/a) = (2^{-j}, 2^{-j}k)$, we can discretize wavelet transform with $W_{\psi}(1/a, b) = W_{\psi}(2^{-j}, 2^{-j}k)$. If we describe wavelet transform as

$$d_k^{(j)} = W_{\psi}(2^{-j}, 2^{-j}k) = 2^j \int_{-\infty}^{\infty} f(x)\overline{\psi(2^j x - k)}dx,$$

and its inverse transform as

$$f(x) \sim \sum_j \sum_k d_k^{(k)} \psi(2^j x - k).$$

2.3 Multiresolution Analysis (MRA)

We can write discrete wavelet inverse transform

$$f(x) = \sum_j g_j(x)$$

by letting

$$g_j(x) = \sum_k d_k^{(k)} \psi(2^j x - k)$$

and

$$f_j(x) = g_{j-1}(x) + g_{j-2}(x) + \dots$$

An integer j is called a level here. Regarding $f(x)$ as $f_0(x)$, the following equation is obtained

$$f_0(x) = g_{-1}(x) + g_{-2}(x) + \dots$$

If a series $\{p_k\}$ is given, the function $\phi(x)$ is called scaling function if it satisfies two-scale equation as

$$\phi(x) = \sum_k p_k \phi(2x - k).$$

When a scaling function is given, we can define a mother wavelet with a series $\{q_k\}$ as

$$\psi(x) = \sum_k q_k \phi(2x - k).$$

We can rewrite $f_j(x)$ recursively as

$$f_j(x) = g_{j-1}(x) + f_{j-1}(x).$$

The function $f_j(x)$ is shown as a linear combination using a scaling function $\phi(x)$ as

$$f_j(x) = \sum_k c_k^{(j)} \phi(2^j x - k).$$

Coefficient $c_k^{(j)}$ is obtained from $c^{(k-1)}$ and $d_k^{(j-1)}$ using decomposition algorithm as follows:

$$c_k^{(j-1)} = \frac{1}{2} \sum_{l \in \mathbf{Z}} g_{2k-l} c_l^{(j)}$$

$$d_k^{(j-1)} = \frac{1}{2} \sum_{l \in \mathbf{Z}} h_{2k-l} c_l^{(j)}$$

Decomposition series $\{g_k\}$ and $\{h_k\}$ is determined from tow-scale equation.

The first wavelet was found by Haar. In this research we used Haar wavelet as a mother wavelet function. Haar wavelet is defined as follows

$$\psi_H(x) = \begin{cases} 1, & 0 \le x \le 1/2 \\ -1, & 1/2 \le x < 1 \\ 0, & \text{otherwise} \end{cases}$$

In Haar case, the whole set of scaling function is obtained by dilation and translation as

$$\psi_{m,n}(t) = 2^{-m/2} \psi(2^{-m} t - n), \quad m, n \in \mathbf{Z}.$$

It is possible to choose the Daubechies orthonormal compactly supported wavelet as a mother wavelet. [1] Multiresolution analysis algorithm in this research is based on the pyramid algorithm by Mallat (1989). [2] This method implements a level J additive decomposition of sequential data.

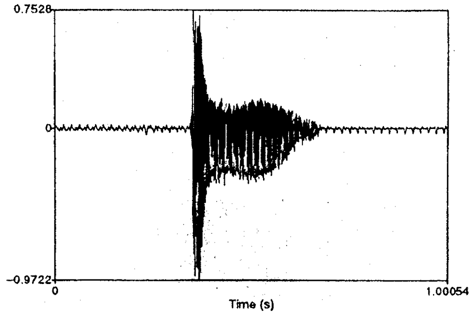

Fig. 1 Raw Wave of Unaspireate *bā* [pa]

2.4 System

I used Sun workstation as a platform. Programming is mainly done by Python language. Statistical analysis and wavelet transform is performed by R-language and `Rwave`, `waveslim` packages (`http://cran.us.r-project.org/`). The speech sounds are recorded to DAT recorder at sample rate of 48kHz in the soundproof room. The speech sound data are stored into WAV files, reducing sample rate to 24kHz. The spectrogram was displayed by praat phonetics software (`http://www.praat.org`).

3. Data Source

The source data is taken from a female native Chinese speaker by pronouncing one syllable two times, who comes from Wuhan, Hunan Province. Two different syllables with the first tone (high flat) are used in this research for comparing consonantal feature of aspiratedness. Speech data were recorded in the soundproof room into DAT recorder and stored into the computer as WAV format files. Speech data are the unaspirated labial sound followed by [a], *bā* [pa], and the apspirted labial sound followed by [a], *pā* [pʰa].

4. Result

4.1 Unaspiratd and Aspirated

The raw data of the unaspirated *bā* [pa] and the aspirated *pā* [pʰa] are shown in figures 1 and 2. The recording sampling rate is 24kHz.

4.2 Spectrogram

Spectrograms are shown in figures 3 and 4. A spike, which spreads in all frequency spectra in a short time interval, is found on the head of syllable in both figures. Compared to the head of the unaspirated labial, there is a noise-like region between the spike and the vowel

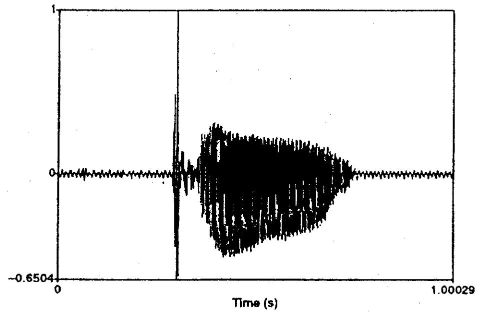

Fig. 2 Raw Wave of Aspirated *pā* [pʰa]

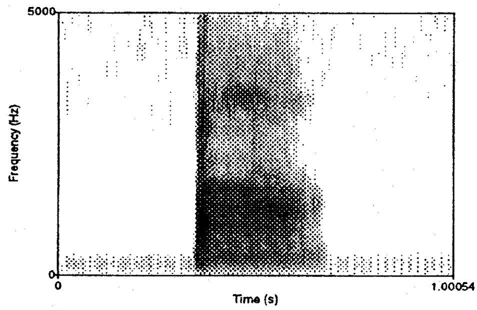

Fig. 3 Spectrogram of *bā* [pa]

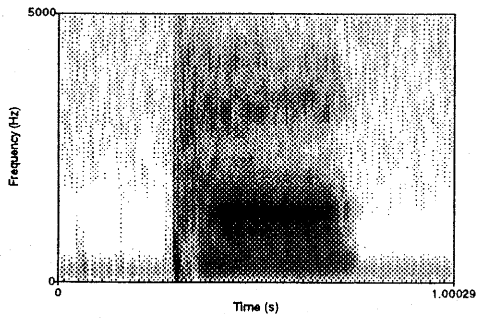

Fig. 4 Spectrogram of *pā* [pʰa]

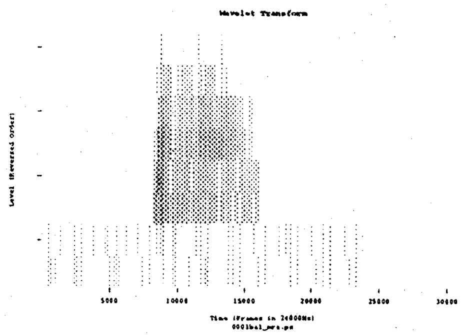

Fig. 5 Multiresolution Analysis of Unaspirated *bā* [pa]

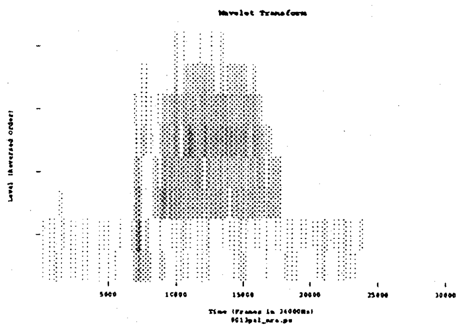

Fig. 6 Multiresolution Analysis of Aspirated *pā* [pʰa]

formant.

4.3 Multiresolution Analysis (MRA)

MRA results are shown in figures 5 and 6. The syllable head regions are shown differently, though the vowel formant regions are not. Spectral blank is seen in the interval between the spike and the vowel formant. The spikes have the different peak of spectra. The spectral concentration of high frequencies are seen in the spike region in the case of unaspirated *bā* [pa]. On the other hand, the concentration of low frequencies are seen in the spike of aspirated *pā* [pʰa].

4.4 Wavelet Coefficient

Comparison with wavelet coefficients in level 7, around 187Hz, for detail analysis indicates that the signal is identified at the spike region of the aspirated *pā* [pʰa], but not at that of the unaspirated. The wavelet coefficients are shown in figures 7 and 8.

5. Discussions

Spectrogram does not clearly indicate the spectral region between spike and vowel formant, but wavelet transforms, MRA in this study, succeeded identification of spectral blank region of aspirated consonant. Furthermore, wavelet transform shows spectral difference between aspirated/unapirated spikes. It can identify the spectral concentration of aspirated consonant at low frequencies.

Though wavelet transform cannot clearly show vowel formant features, it can detect consonantal features better than spectrograph. Usefulness of wavelet transform on speech consonantal analysis is proved in this study.

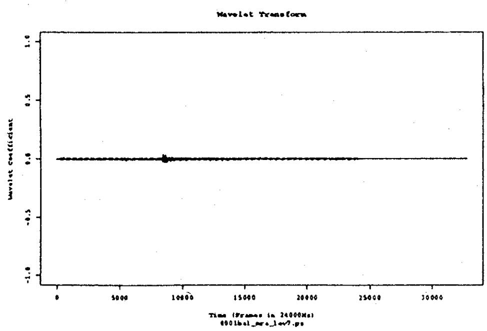

Fig. 7 Wavelet Coefficient in Level 7 of Unaspirated *bā* [pa]

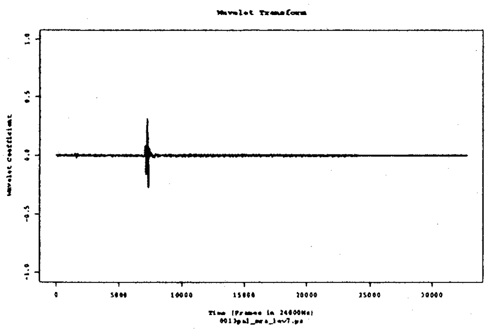

Fig. 8 Wavelet Coefficient in Level 7 of Aspirated *pā* [p^ha]

References

[1] I. DAUBECHIES, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, 1992.

[2] S. G. MALLAT, *A theory for multiresolution signal decomposition: the wavelet representation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11 (1989), pp. 674–693.

[3] J. TIAN AND J. MIWA, *A method of discrimination and scoring of aspirated and unaspirated sounds for chinese spoken language learning*, IEICE Transactions D-II, J85-D-II (2002), pp. 842–850.