The Fifteenth International Symposium on Artificial Life and Robotics 2010 (AROB 15th ’10),
B-Con Plaza, Beppu,Oita, Japan, February 4-6, 2010

The analysis of Japanese voice sound by using Real-Time Spectral Analysis

Hideto Nakatsuji [*], Sigeru Omatu [**]
*Computer and Systems Sciences, Department of Engineering, Osaka Prefecture University, 1-1 Gakuen-cho, Nakaku, Sakai City, 599-8531
**Computer and Systems Sciences, Graduate School of Engineering, Osaka Prefecture University, 1-1 Gakuen-cho, Nakaku Sakai City, 599-8531
nakatuji@sig.cs.osakafu-u.ac.jp,   omatu@cs.osakafu-u.ac.jp

Abstract

A voice signal processing requires a real-time processing. In consideration of the real-time processing, a new method which updates the spectrum by using an input data has been proposed. In this method, the analysis in time-frequency domain is executed as well as wavelet transform. Vowel /i/ and /e/ consist of fundamental, harmonics and high frequency waves. These high frequency waves determine the sound of /i/ and /e/. These high frequency waves might exist also in /a/ and /o/. In this paper, we analyze these high frequency waves by using new analyzing method, and we show the constitution of high frequency waves. Since a waveform of consonant does not repeat, the analysis can not be performed like the analysis of the high frequency wave of vowel. In consonant, we show the fluctuation of the spectrum for the time progress.

Keywords: Decomposition wave and reconstruction wave, Time-frequency domain

1 Introduction

In recent years, wavelet transform has attracted attention as an analysis in time-frequency domain. Here newly, a new analytical method with the same function as wavelet transform was presented by authors [2]-[4]. In this method, processing in search of spectrum is performed by using one input data. Since all processing are perfectly independent, complete parallel processing is possible, and so if parallel processing computer will be brought to realization, real-time processing is realized. We call this analytical method Real-Time Spectral Analysis Method. In Japanese, there are five vowels, and this is few in comparison with other language. Japanese vowel consists of fundamental, harmonics and high frequency waves. In Japanese vowel /i/ and /e/ , High frequency waves decide the sound /i/ and /e/. In Japanese vowel /a/ and /o/, it is not always true that the high frequency waves exist and high frequency waves do not decide the sound /a/ and /o/. High frequency waves are audible in each vowel sound. Here, by applying the method of decomposition and reconstruction of Real-Time Spectral Analysis to the Japanese vowel sound, the sound signal is decomposed to a fundamental, some harmonics and high frequency waves. By using these waves, we made experiment of hearing and we show how the vowel sound consists of these waves. In consonant, we show the fluctuation of the spectrum with time.

2 Analysis theory

The algorithm uses the inner products of multiple period sine, cosine waves (below referred to as cutting out waves)and a signal. T_s is the sampling interval. Let f_j denote the established frequency of the cutting out wave (angular frequency $\lambda_j = 2\pi f_j$). Here $j(=1,\cdots,n)$ is number of cutting out wave and n is the number of the cutting out waves. There are T periods (T is a natural number), here T is the number of periods. The length of the cutting out wave is $q_j(=T/f_j)$ and N_j denote the number of data included in the cutting out wave. Then the cutting out wave can be expressed as follows:

$$\begin{aligned} &s_j(l) = \sin(\lambda_j l T_s) \quad c_j(l) = \cos(\lambda_j l T_s) \\ &s_j(-l) = -s_j(l) \quad c_j(-l) = c_j(l) \\ &s_j(k+N_j) = s_j(k) \quad c_j(k+N_j) = c_j(k) \\ &k\text{ ; natural number} \quad l = 1,\cdots,N_j \end{aligned} \tag{1}$$

Here we consider only sine waves as cutting out wave because the expansion for cosine waves is similar to the case of sine waves. A signal at time kT_s is $x(k)$. By using the signal which dated back to the past, the inner product of cutting out wave and signal at time kT_s can be written as follows:

$$Y_s^k(j) = \frac{2}{N_j}\sum_{l=1}^{N_j} x(k-l)s_j(l-k) \tag{2}$$

The inner product at time $(k+1)T_s$ is

$$\begin{aligned} Y_s^{k+1}(j) &= \frac{2}{N_j}\sum_{l=1}^{N_j} x(k+1-l)s_j(l-k-1) \\ &= \frac{2}{N_j}x(k)s_j(-k) + Y_s^k(j) - \frac{2}{N_j}x(k-N_j)s_j(N_j-k) \end{aligned} \tag{3}$$

From Eq.(3), the inner product at time $(k+1)T_s$ can be obtained by adding the product of the input data at time kT_s and the cutting out wave to the inner product at time kT_s, and subtracting the third term. Since the second and third term have already been calculated, the necessary calculation in Eq.(3) is a multiplication in the first term. If $k = N_j$, then third term on the right side of Eq.(3) is $s_j(0)$, but this can be found $s_j(N_j)$ by using Eq.(1). In the case of a cosine cutting out wave, the inner product $Y_c^k(j)$ is found using a similar way. By using these inner products, the following is calculated:

$$Y_{out}^{k}(j) = \sqrt{Y_s^{k}(j)^2 + Y_c^{k}(j)^2} \tag{4}$$

The unit that outputs $Y_{out}^k(j)$ detects input signal components of certain frequency (close to the established frequency f_j), below, this unit is called an auditory cell, and $Y_{out}^k(j)$ is spectrum. Since there are two equations of inner product in an auditory cell, the multiplication number of times that is necessary for update of an inner product is twice. Since number of auditory cells is n, total multiplication number is $2n$. Auditory cells are independent each other, and two equations in an auditory cell are also independent. If parallel processing computer will be brought to realization, calculation number that need to calculates spectrums by Eq.(4) is 2 multiplications and one square root calculation. Even if number of auditory cells becomes large, the calculation number of times does not change. By using the inner products expressed by discrete system, we show below a procedure to get decomposition waves.

Step1 Calculate the inner products and $Y_s^k(j)$ and $Y_c^k(j)$ at time kT_s

Step2 multiply $Y_s^k(j), Y_c^k(j)$ by $-s_j(k), c_j$ and the following is calculated:

$$Y_{ss}^{k}(j) = -Y_s^{k}(j)s_j(k), \quad Y_{sc}^{k}(j) = Y_s^{k}(j)c_j(k)$$

$$Y_{cc}^{k}(j) = Y_c^{k}(j)c_j(k), \quad Y_{cs}^{k}(j) = -Y_c^{k}(j)s_j(k)$$

Step3 By using $Y_{ss}^k(j), Y_{sc}^k(j), Y_{cc}^k(j)$ and $Y_{cs}^k(j)$, The following is calculated:

$$Y_{cout}^{k}(j) = Y_{ss}^{k}(j) + Y_{cc}^{k}(j), \quad Y_{sout}^{k}(j) = Y_{sc}^{k}(j) - Y_{cs}^{k}(j)$$

Step4 Changing k into $k+1$, and returns to Step 1 and repeats below. $Y_{cout}^k(j)$ is the decomposition wave that appears on the auditory cell.

3 The analysis of Japanese vowel sound

I acquired the sound data of seven men. Sampling frequency is 44.1kHz, 16 bits, monaural. Number of periods T is 16, and constant. Now assume 261 auditory cells with the established frequency increasing from 100Hz to about 17kHz at a rate of about 2%.

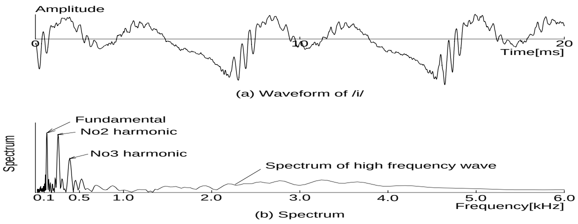

Fig. 1. Waveform and spectrum of /i/

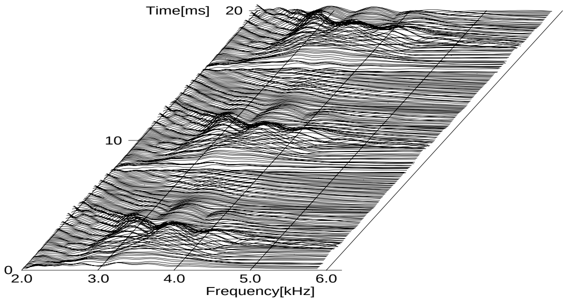

Fig. 2. Fluctuation of spectrum of high frequency wave with time

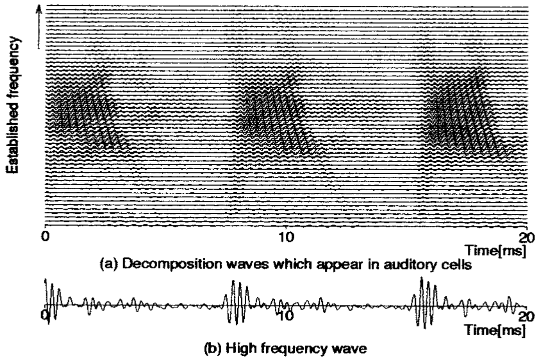

Fig. 3. Decomposition waves and high frequency wave

3.1 High frequency wave constituting a vowel sound

There are five vowels in Japanese, and the vowel generally consists of fundamental, harmonics and some high frequency waves. High frequency waves exist in higher frequency than 2kHz. The high frequency waves exist by all means in vowel /i/ and /e/. In addition, there are not the harmonics of higher frequency than 500Hz. A waveform and a spectrum of /i/ are showed in Fig.1. In Fig.1, there are fundamental, two harmonics and one high frequency wave. When the wave made by adding fundamental and harmonics is played back, we do not hear it with /i/, but the wave produced by adding high frequency wave to that wave is heard with /i/. When the high frequency wave only is played back, we hear that sound with /i/. In /i/, the existence of this wave decides /i/, and /e/ is similar to /i/. In /i/, the spectrum fluctuating with time is showed in Fig.2. In Fig.2, the spectrum repeats itself at fundamental frequency. The decomposition waves are calculated by the method shown with Section 2, and the decomposition waves is shown in Fig.3. In Fig.2 and Fig.3, time indicates until 20ms.

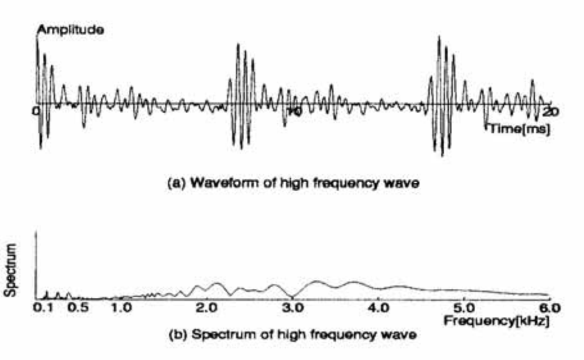

Fig. 4. Waveform and spectrum of high frequency wave

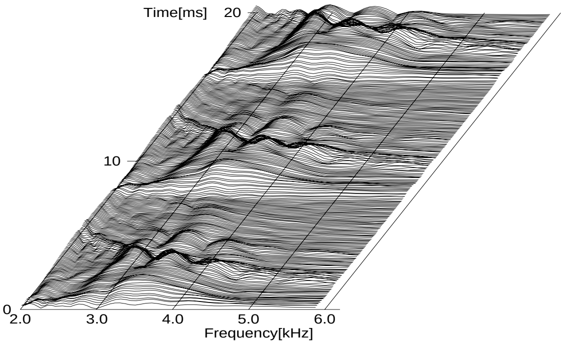

Fig. 5. Fluctuation of spectrum with time

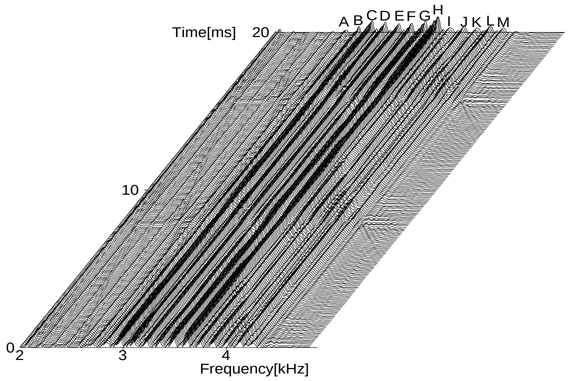

Fig. 6. Fluctuation of spectrum with time in case that T is 100

Fig.3(a) shows the decomposition waves of high frequency wave. In Fig.3(a), vertical axis is the established frequency, and in the upper part, frequency becomes high. Fig.3(b) is the wave that are made by adding all decomposition waves of high frequency, and this is the reconstruction wave of high frequency wave. When this wave is reproduced, we hear it with /i/. Fig.4 shows the spectrum obtained by analyzing this high frequency wave again. Fig.4(a) and Fig.3(b) are the same. Although fundamental and harmonics are small, they exist similarly in Fig.4(b), and high frequency wave exists too. Spectrum of this high frequency wave shows in Fig.5. Spectrum in Fig.5 is almost the same as spectrum in Fig.2. A high frequency wave corresponding to the spectrum in Fig.5 is made, and we hear it with /i/. In this way, the high frequency wave consists of fundamental, harmonics and high frequency wave, and this high frequency wave also consists of those waves.

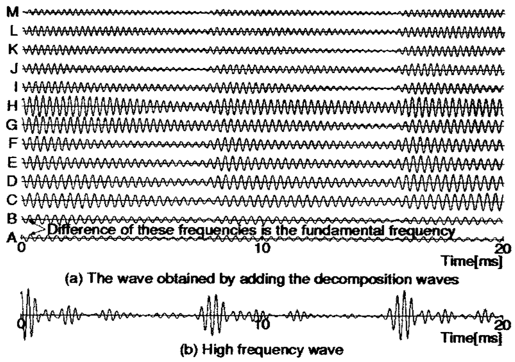

Fig. 7. Elements of high frequency wave and produced wave by adding those waves

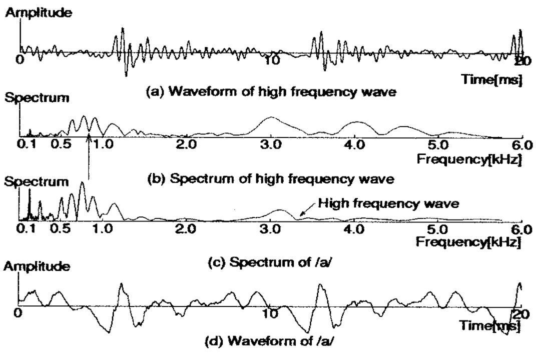

Fig. 8. waveform of /a/ and high frequency wave and that spectrums

3.2 The separation of the spectrum

By using number of periods T and the established frequency f , the length q of the cutting out wave is shown with T/f . The waves that length for $T+1$ periods and $T-1$ periods become q can not be detected by this auditory cells. The detection width of the auditory cell is designated as B . The following relationship $B=2f/T$ or $Bf/T=2$ is formed. That is, the area made from frequency width B and time length T/f becomes two and constant. When T is increased, B that shows the extension of the spectrum becomes small, and so it is possible to separate the spectrum that has extended in the direction of the frequency axis. In Fig.2, the spectrum repeats itself, and so we make T large and try separation of spectrum. Fluctuation of spectrum against the passage of time in case that T is 100 is shown in Fig.6. Fig.6 shows that a spectrum is separated into 13 from A to M. The decomposition waves from A to M and the wave made by adding all the decomposition waves is shown in Fig.7. Fig.7(a) is the decomposition waves, and Fig7.(b) is the reconstruction wave, and this wave is the high frequency wave. The frequency interval of the decomposition waves is fundamental frequency. Since this frequency interval is very smaller than 3kHz or 4kHz, when these waves are added, complicated large beat phenomenon occur. In this

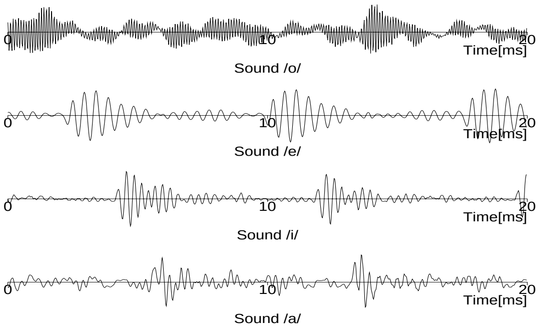

Fig. 9. Waveforms of high frequency waves of vowel /a/,/i/,/e/and/o/

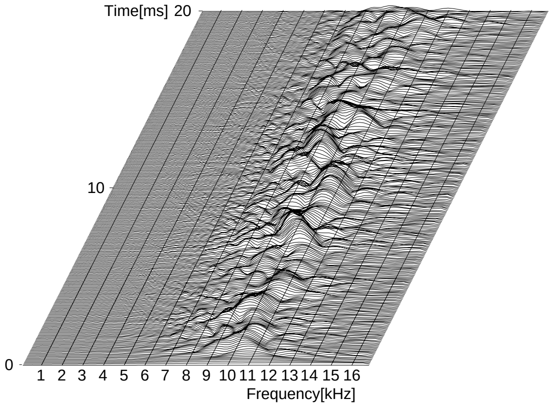

Fig.10. Fluctuation of spectrum of /s/of /sa/ with time

way, the high frequency wave becomes the wave which fluctuates largely. The spectrum in Fig.2 and in Fig.6 essentially the same. In /i/, one or more high frequency waves exist by all means, and we hear the wave with /i/. The existence of this wave determines /i/. Vowel /e/ is also similar to /i/. It is not always true that the high frequency waves exist in /a/ and /o/. In the case that the wave exists, we hear it with each vowel /a/ or /o/. A waveform of /a/ and that spectrum and high frequency wave and that spectrum are shown in Fig.8. In /a/, a wave produced by adding fundamental and harmonics is heard with /a/. Fig.8(c) is the spectrum of /a/ and Fig.8(b) is the spectrum of the high frequency wave, and in both spectrums, there are similar fundamental and harmonics, and so we hear high frequency wave with /a/. Vowel /o/ is also same as /a/. In Fig.6 and Fig.7, it was shown that the spectrum of high frequency wave was separated by enlarging$T$. This is the same also in /a/, /o/, and /e/. In /a/, high frequency wave exists in five persons among seven persons, and in /o/, high frequency wave exists in three persons among seven persons. In /u/, there are fundamental, harmonics and high frequency wave, but it is not always true that the high frequency waves exist. High frequency wave exists in four persons among seven persons. When high frequency wave is played pack, we do not hear it with /u/. It seems not to be decided how you hear it. Here we show the examples of high frequency waves in Japanese vowel /a/,/i/,/e/ and /o/ in Fig.9. Next, number of periods is 10, and the fluctuation of the spectrum of consonant /s/ in /sa/ with time is shown in Fig.10. In Fig.10, irregular concavo-convex shape appears around 10kHz. This is consonant. At 10kHz, the length of one period is 0.1ms. Since number of periods T is 10, the length of the cutting out wave (equivalent to window's width) is 1ms $(=0.1ms\times10)$. The length of the cutting out wave changes depending on the frequency. As a result, the spectrum which fluctuates with time can be expressed.

4. Conclusion

The vowel generally consists of fundamental, harmonics and high frequency wave. High frequency waves exist in /i/ and /e/ by all means, and the existence of this wave determines /i/, and /e/, but it is always true that the high frequency waves exist in /a/ and /o/. We hear the high frequency wave with each vowel. In /i/./e/,/a/ and /o/, since the spectrum of high frequency wave repeats itself, The spectrum is able to separate to some spectral components by enlarging T. Although the appearance is different in spectrum in case of small T and spectrum in case of large T, both of them are essentially same. Furthermore, the decomposition waves according to these spectral components are obtained, and the frequency interval of the decomposition waves is fundamental frequency. When these waves are added, complicated large beat phenomenon occur. In this way, the high frequency wave becomes the wave which fluctuates largely. In /u/, although the high frequency wave may exist also, it does not decide how you hear it. Since there is no repetition at a consonant, separation of a spectrum which was performed with the vowel cannot be performed. But the fluctuation of the spectrum against the passage of time can be shown. Here, we show the Japanese voice sound, but in other languages, it is expected that a different result will be obtained. For example, there are 5 vowel in Japanese, but in Korean, there are 10 vowell, and in addition, how about the consonant. I am very interested in these analysis and comparison

References

[1] O. Rioul and M. Vetterli, "Wavelet and Signal Processing," IEEE SP Magazine Vol.8, No4, pp.14-39 (1991).

[2] Nakatsuji, and S. Omatu, " Real-Time Spectrl Analysis," Electrical Engineering in Japan, Wiley InterScience , Vol. 161, No. 1, pp. 43-50, 2007

[3] H. Nakatsuji and S. Omatu, "Decomposition and reconsrtruction of signal in Real-time spectral analysis," Electrical Engineering in Japan, Wiley InterScience , Vol. 91, No.11, pp. 1123-1130, 2008

[4] H. Nakatsuji and S. Omatu, "Decomposition and reconsrtruction of signal in High-Speed Processing Method of Real-time spectral analysis," IEEJ Trans.EIS, Vol.128, No. 7, pp. 1168-1175, 2008 (in Japanese).