Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

Noisy Speech Recognition using Wavelet Transform and Weighting Coefficients for a Specific Level

Yoichi MIDORIKAWA , Yuta MURAOKA and Masanori AKITA

Faculty of Engineering, Oita University, 700 Dannoharu, Oita, 870-1192, JAPAN

PACS: 43.72.Ne

ABSTRACT

The purpose of this study is to propose an improvement of the speech recognition under the noisy environment. Speech recognition systems have a weak point, they work best in a noiseless condition, but they work poor performance under the noise condition. To analyse noise problem, many people have used fourier analysis. But the fourier analysis reveals only the frequency information. The general noise filters reduce specific frequency noise and signal. It is difficult to remove the only noise. To overcome this difficulty, we apply wavelet transform to speech recognition under the noisy environments using cepstral analysis. We applied the method of signal decomposition and synthesis by wavelet transform and weighting coefficients for a specific level and cepstral analysis. This method is applied to the speech recognition. As a result, speech recognition rate is improved by this method.

INTRODUCTION

The purpose of this study is to propose an improvement of the speech recognition under the noisy environment. Speech is the most natural form of human communication. And speech recognition has made it possible for computers to understand human voice commands and human languages. Speech recognition is one of the very important technologies. Automatic speech recognition systems are most effective in noiseless environments. However, these speech recognition systems have a weak point, they work best in a noiseless condition, but they work poor performance under the noise condition. If the data are polluted with noise, these speech recognitions are extremely difficult. For the noise reduction of signals, there are filters and spectral subtraction method, and so on. However, there is some limitation in case that the quality of the signal is poor. Therefore, we took notice of this point. This study improves a speech recognition method under the noisy environment by using wavelet transform and weighting coefficients and cepstral analysis. To analyse noise problem, many people have used fourier analysis. But the fourier analysis reveals only the frequency information. The general noise filters reduce specific frequency noise and signal. It is difficult to remove the only noise. To overcome this difficulty, we apply wavelet transform to speech recognition under the noisy environments using cepstral analysis. The wavelet transform is widely used for wave and image analysis [1- 8]. We applied the method of signal decomposition and synthesis by wavelet transform and weighting coefficients for a specific level and cepstral analysis. This method is applied to the speech recognition. The speech recognition experiments by Japanese noisy digits are performed two kinds of colored noise are added to the original clear speech data for making noisy data. As a result, speech recognition rate is improved by this method. It was shown that the wavelet transform and weighting coefficients was one of the promising methodologies for filter.

METHOD OF USING WAVELET TRANSFORM AND WEIGHTING COEFFICIENTS FOR A SPECIFIC LEVEL

For the noise reduction of signals, there are a noise filters. The general noise filters reduce specific frequency noise and signal. It is difficult to remove the only noise. To overcome this difficulty, we applied the method of signal decomposition by wavelet analysis. To analyze noise problem, many people have used fourier analysis. Fourier spectrum reveals only the frequency information. Therefore, it is difficult to obtain the noise waveforms from the fourier spectrum. To overcome this difficulty, we employed the wavelet analysis technique. Wavelet analysis makes it possible to get the time as well as frequency domain information.
The wavelet transformation is given by

$$\mathbf{S} = \mathbf{W} \cdot \mathbf{X} \tag{1}$$

where **S** is wavelet spectrum matrix, **W** is wavelet matrix [1-5] , **X** is signal data.
Method of using wavelet transform and weigting coefficients for a specific level is given by

$$s'_i = \begin{cases} ks_i & \left(N/2^M < i \le N/2^{M-1}\right) \\ s_i & (otherwise) \end{cases} \tag{2}$$

S_i is element of matrix **S**. N is a length of signal data. M is the level number. k is weigting coefficient.
Equation (3) is inverse wavelet transformation.

$$\mathbf{X'} = \mathbf{W}^{\mathrm{T}} \cdot \mathbf{S'} \quad (3)$$

S' is the modification of wavelet spectrum matrix by Equation (2). **X'** is modified signal data matrix. The modified signal data are analyzed into cepstral coefficients using improved cepstral method [9]. And speech recognition experiments are used the cepstral coefficients.

Figure 1 shows that definition of each level. Figure 2 shows that exsample of wavelet spectrum using weighting coefficient *k*=0 for level 1 (n1). Figure 3 shows that exsample of wavelet spectrum using weighting coefficient *k*=0.5 for level 2 (n2).

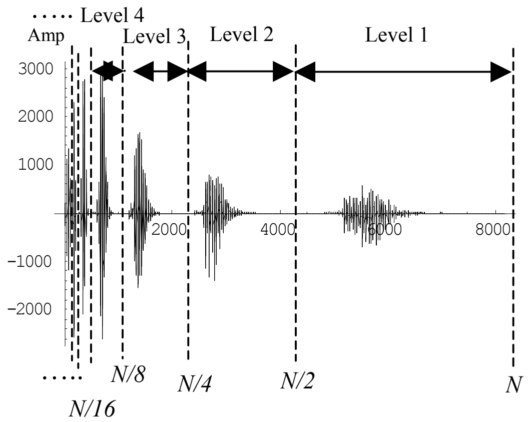


Figure 1. Definition of each level

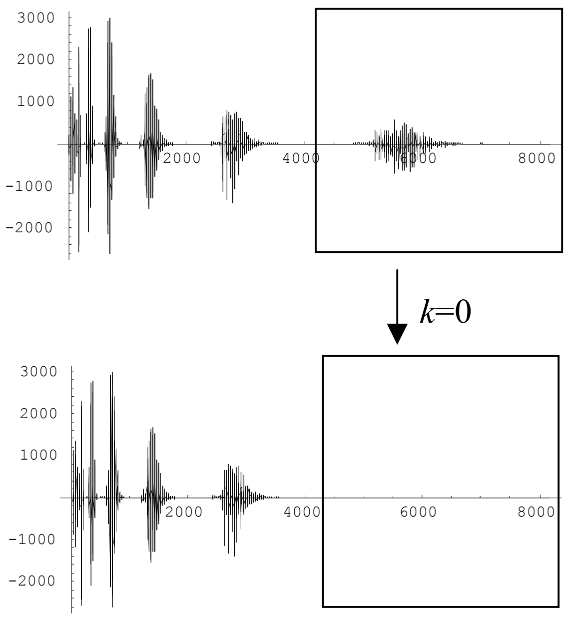


Figure 2. Exsample of wavelet spectrum using weighting coefficient k=0 for Level 1. (n1)

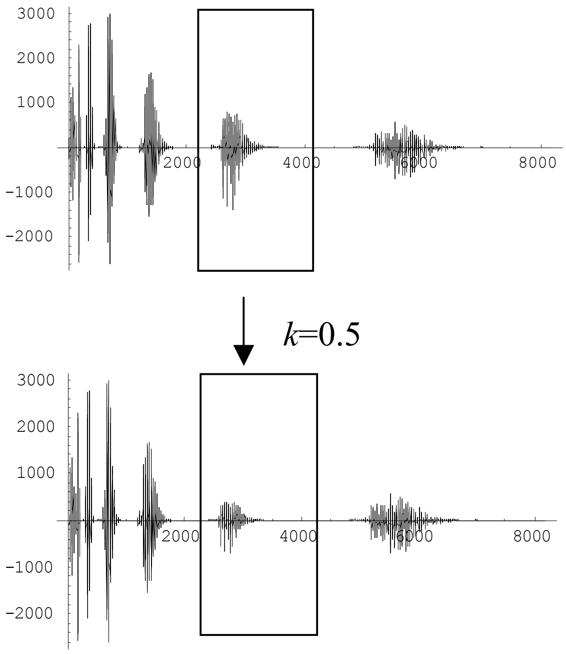


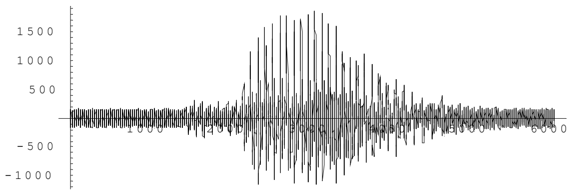
Figure 3. Exsample of wavelet spectrum using weighting coefficient k=0.5 for Level 2. (n2)

Figure 4 shows that exsample of the method of using wavelet transform and weighting coefficients for a specific level. Figure 4 (a) shows that original speech data of Japanese noisy digits. The parameter *N* must be the integer and the power of 2. This exsample data is not the power of 2. Therefore, speech data added data of zero shown in Figure 4 (b). We computed a wavelet spectrum. Figure 4(c) shows the wavelet spectrum. Figure 4(d) shows that exsample of modified wavelet spectrum using weighting coefficient *k*=0 for level 1 (n1). To recover the speech data, we computed the inverse wavelet transform shown in Figure 4 (e). Figure 4(f) shows that restoration of speech data length.

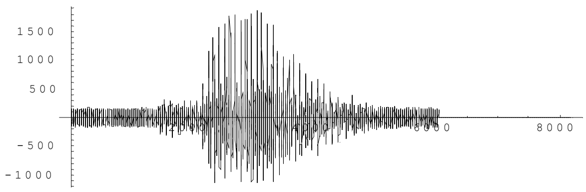
EXPERIMENTAL RESULTS

This method is applied to the speech recognition. The speech recognition experiments by Japanese noisy digits are performed. 240 Japanese digits uttered by 8 male speakers are used. Each speaker uttered digits for three times. Two kinds of colored noise are added to the original clear speech data for making noisy data. These are expressed as Pink noise and Automobile noise.

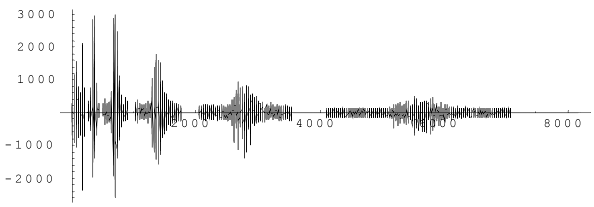
The reference patterns are obtained from single speaker and the data of other speakers are recognized. In the experiments, reference patterns are obtained from clear speech data. The dynamic time warping matching is used for the pattern matching. The wavelet base function is the haar wavelet.



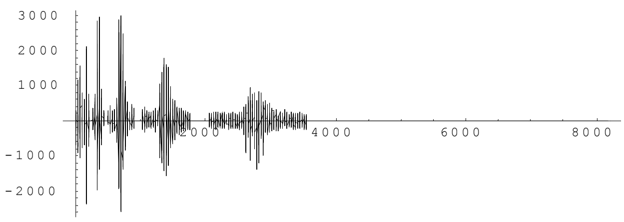
(a) Speech data.



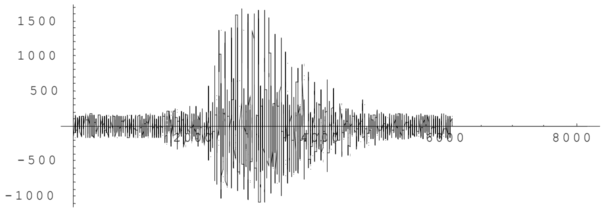
(b) Speech data added data of zero.



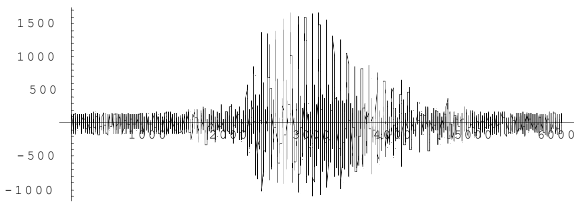
(c) wavelet spectrum.



(d) Exsample of modified wavelet spectrum using weighting coefficient k=0 for Level 1



(e) Recovered speech data by inverse wavelet transpose.



(f) Restoration of speech data length.

Figure 4. Exsample of the method of using wavelet transform and weighting coefficients for a specific level. (Japanese /rei/ with Pink 10dB)

Table 1 show that the recognition results by conventional method. This recogniton results are standard results in this paper. Experimental results of the speech recognition using the new methods are shown in Figures 5-11.

Figure 5 show that the recognition results of wavelet spectrum using weighting coefficient for level 1 (n1). Figure 6 show that the recognition results of wavelet spectrum using weighting coefficient for level 2 (n2). Figure 7 show that the recognition result of wavelet spectrum using weighting coefficient for level 3 (n3). Figure 8 show that the recognition result of wavelet spectrum using weighting coefficient for level 4 (n4). Figure 9 show that the recognition result of wavelet spectrum using weighting coefficient for level 5 (n5). Figure 10 show that the recognition result of wavelet spectrum using weighting coefficient for level 6 (n6). Figure 11 show that that recognition result of wavelet spectrum using weighting coefficient for level 7 (n7).

Table 1. Recognition results by standard method.

Noise type	Recognition rate [%]
Clean	95.14
Pink0dB	15.85
Pink10dB	28.31
Auto0dB	24.29
Auto10dB	54.03

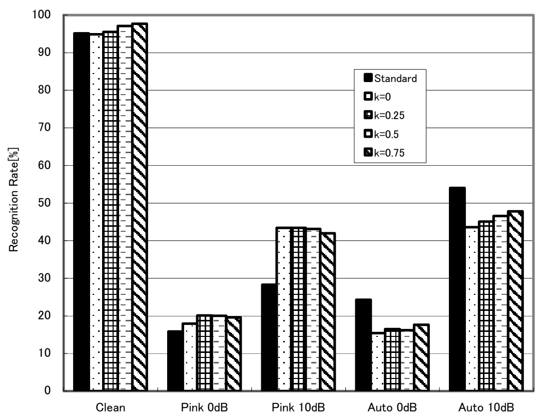


Figure 5. Recognination results (Level 1, n1)

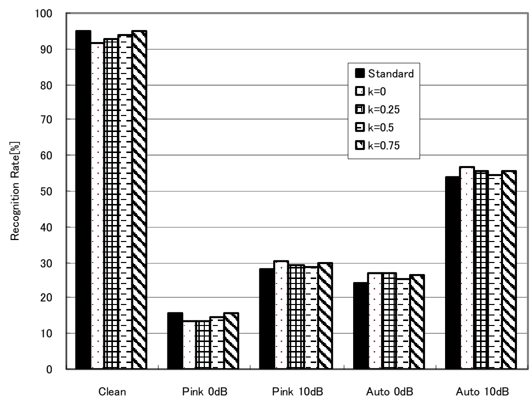


Figure 6. Recognination results (Level 2, n2)

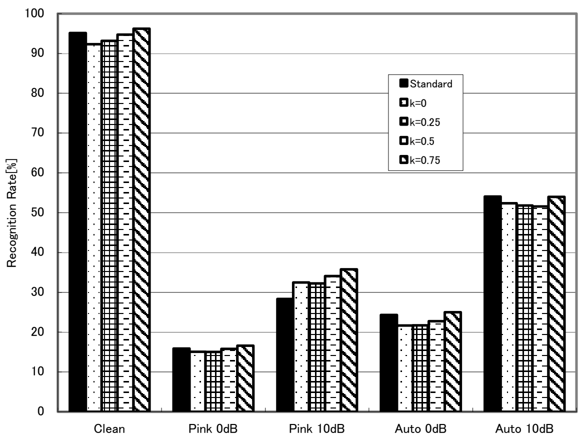


Figure 7. Recognination results (Level 3, n3)

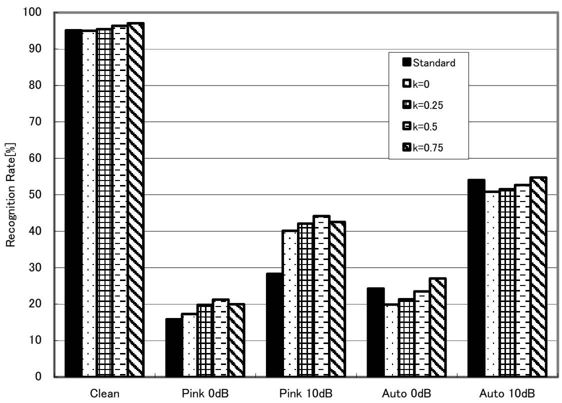


Figure 8. Recognination results (Level 4, n4)

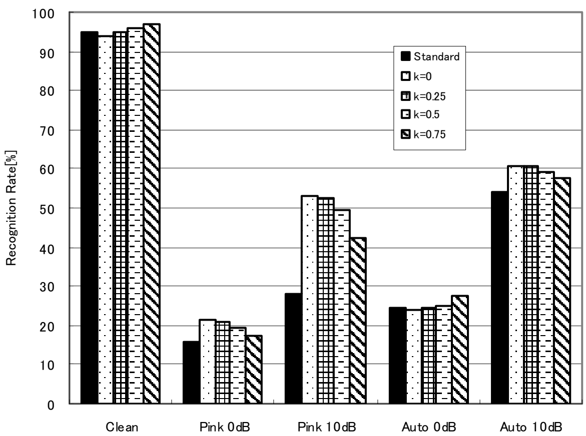


Figure 9. Recognination results (Level 5,n5)

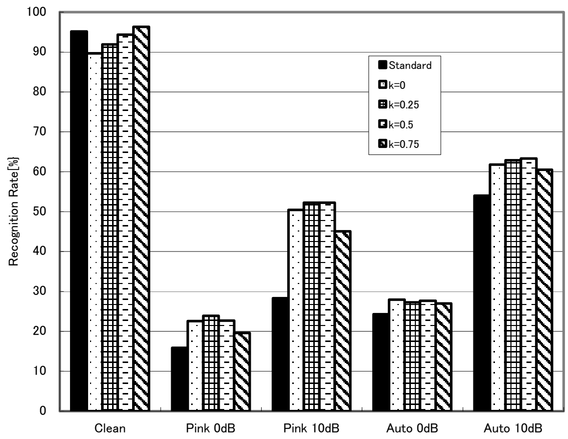


Figure 10. Recognination results (Level 6,n6)

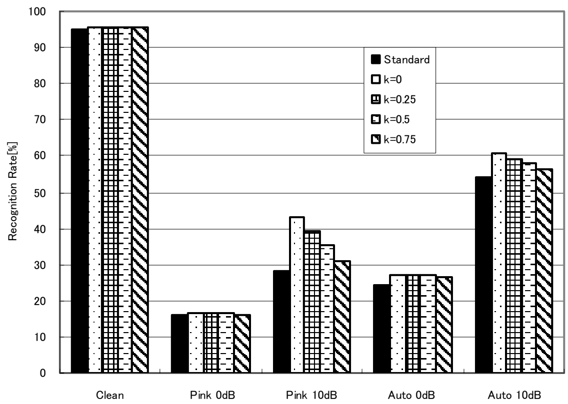


Figure 11. Recognination results (Level 7,n7)

In the experimental results, recognition results of Pink S/N10dB are improved about 10 to 20%. But, recognition results of Automobile S/N0dB are not improved. It is necessary to improve this method under the Automobile noisy environment.

Table 2 show that the best recognition results of methods by wavelet transform and weighting coefficients for a specific level. In particular, recognition results of n5 or n6 and k=0 or k=0.25 are improved.

Table 3 shows that the parameter k is the best recognition results of methods by wavelet transform and weighting coefficients for each level.

Figure 12 and Figure 13 show that the exsample of spectral envelopes of /i/ parts of Japanese word /ichi/ (one). Figure 12 shows that the spectral envelops of /i/ parts of Japanese word /ichi/ (one) by conventional method. Figure 13 shows that the modified spectral envelops of /i/ parts of Japanese word /ichi/ (one) by wavelet transform and weighting coefficients for level 5 (n5).

Figure 13 compares advantageously with Figure 12. In Figure 13, the spectral envelops of clearn and otherwise are somewhat analogous.

Table 2. The best recognition results of methods by wavelet transform and weighting coefficients for a specific level.

Noise	Weighting pattern	Recognition rate [%]
Clean	n1(k=0.75)	97.70
Pink0dB	n6(k=0.25)	23.91
Pink10dB	n5(k=0)	53.04
Auto0dB	n6(k=0)	27.94
Auto10dB	n6(k=0.5)	62.29

Table 3. The parameter *k* is the best recognition results of methods by wavelet transform and weighting coefficients for each level.

Level	Pink	Automobile	Average
1	0.25	1	0.75
2	0.75	0	0.75
3	0.75	0.75	0.75
4	0.5	0.75	0.75
5	0	0.75	0.25
6	0.25	0.5	0.5
7	0	0	0

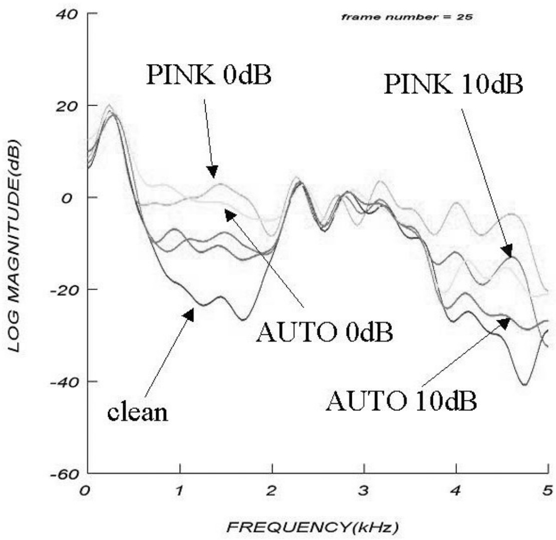


Figure 12. Spectral envelops. (/ichi/, standard)

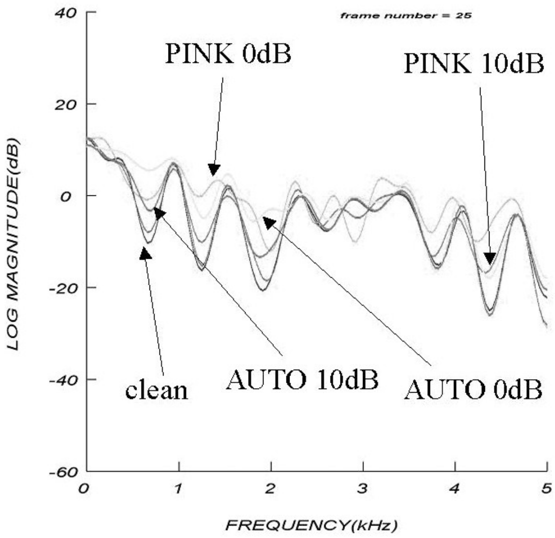


Figure 13. Spectral envelops. (/ichi/, n5, *k*=0)

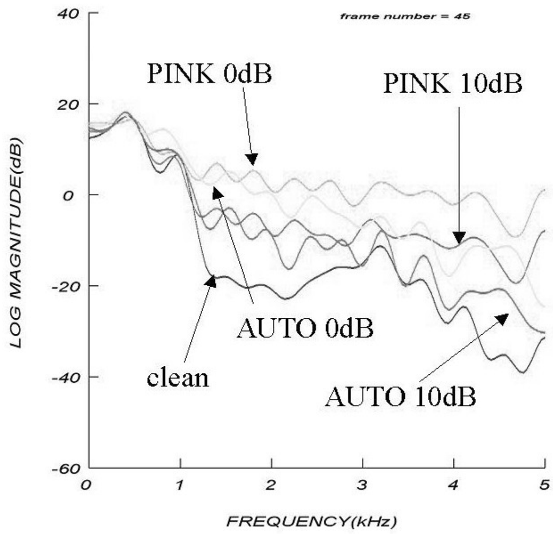


Figure 14. Spectral envelops. (/go/, standard)

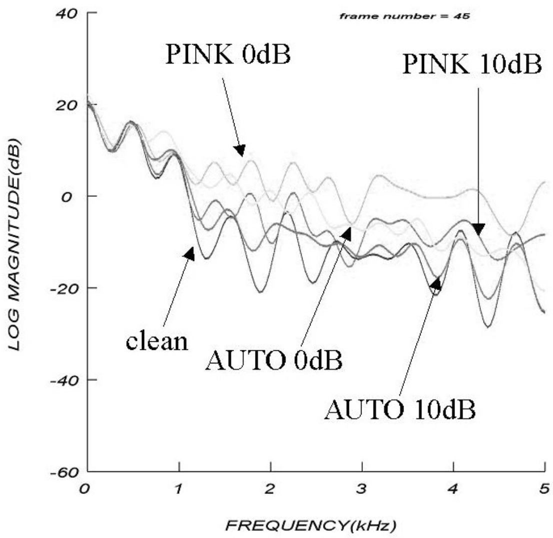


Figure 15. Spectral envelops. (/go/, n5, *k*=0)

Figure 14 and Figure 15 show that the exsample of spectral envelopes of /o/ parts of Japanese word /go/ (five). Figure 14 shows that the spectral envelops of /o/ parts of Japanese word /go/ (five) by conventional method. Figure 15 shows that the modified spectral envelops of /o/ parts of Japanese word /go/ (five) by wavelet transform and weighting coefficients for level 5 (n5).

Figure 15 compares advantageously with Figure 14. In Figure 15, the spectral envelops of clearn and otherwise are somewhat analogous.

Figure 16 and Figure 17 show that the exsample of spectral envelopes of /a/ parts of Japanese word /nana/ (seven). Figure 16 shows that the spectral envelops of /a/ parts of Japanese word /nana/ (seven) by conventional method. Figure 17 shows that the modified spectral envelops of /a/ parts of Japanese word /nana/ (seven) by wavelet transform and weighting coefficients for level 5 (n5).

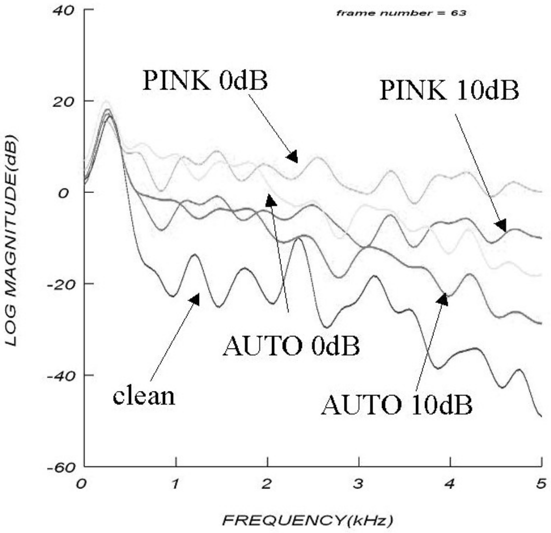


Figure 18. Spectral envelops. (/kyu/, standard)

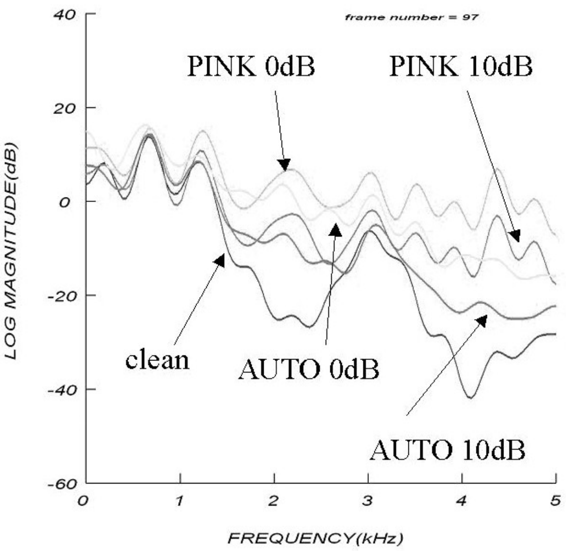


Figure 16. Spectral envelops. (/nana/, standard)

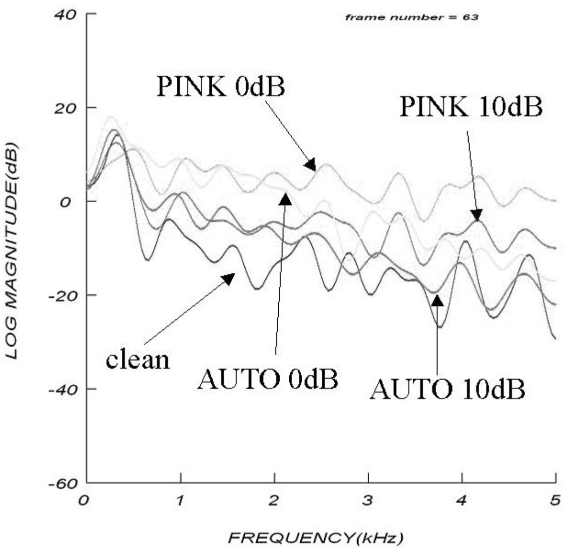


Figure 19. Spectral envelops. (/kyu/, n5, *k*=0)

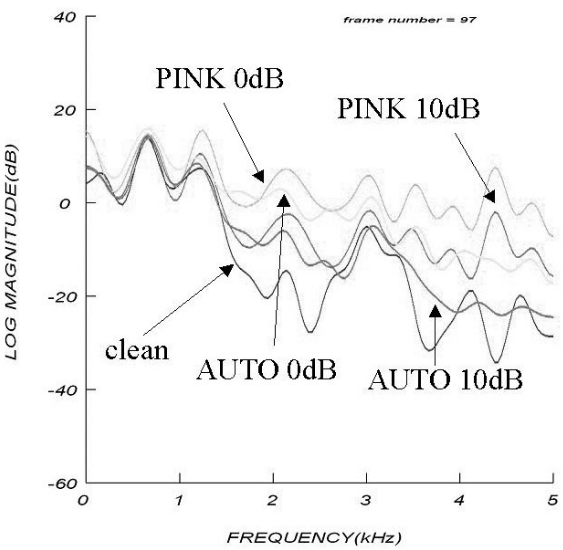


Figure 17. Spectral envelops. (/ichi/, n5, *k*=0)

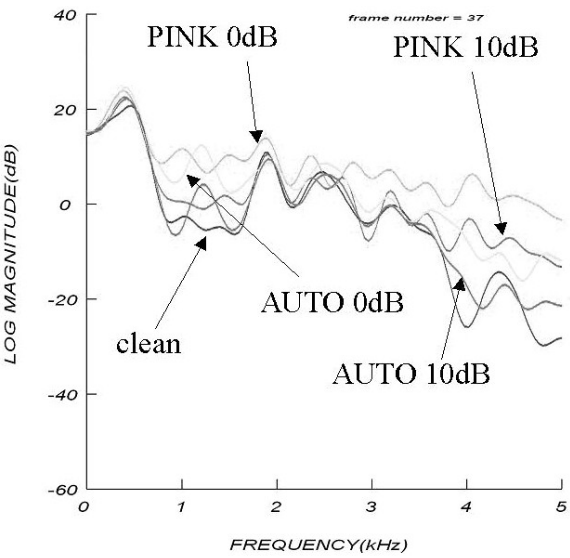


Figure 20. Spectral envelops. (/rei/, standard)

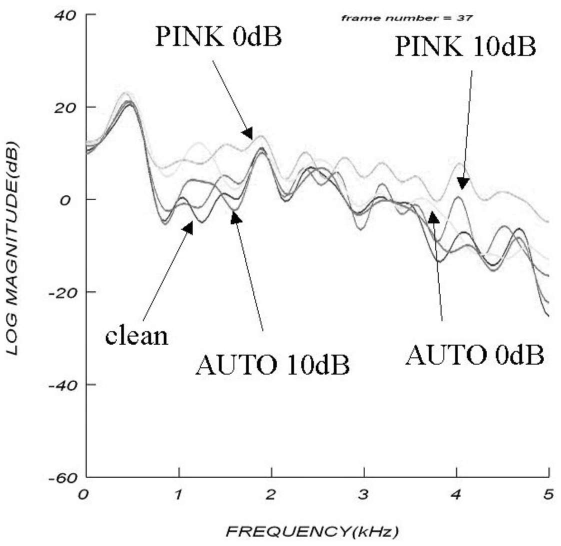


Figure 21. Spectral envelops. (/rei/, n5, *k*=0)

Figure 17 compares advantageously with Figure 16. In Figure 17, the spectral envelops of clearn and otherwise are somewhat analogous.

Figure 18 and Figure 19 show that the exsample of spectral envelopes of /u/ parts of Japanese word /kyu/ (nine). Figure 19 shows that the spectral envelops of /u/ parts of Japanese word /kyu/ (nine) by conventional method. Figure 19 shows that the modified spectral envelops of /u/ parts of Japanese word /kyu/ (nine) by wavelet transform and weighting coefficients for level 5 (n5).

Figure 19 compares advantageously with Figure 18. In Figure 19, the spectral envelops of clearn and otherwise are somewhat analogous.

Figure 20 and Figure 21 show that the exsample of spectral envelopes of /e/ parts of Japanese word /rei/ (zero). Figure 20 shows that the spectral envelops of /e/ parts of Japanese word /rei/ (zero) by conventional method. Figure 21 shows that the modified spectral envelops of /e/ parts of Japanese word /rei/ (zero) by wavelet transform and weighting coefficients for level 5 (n5).

Figure 21 compares advantageously with Figure 20. In Figure 21, the spectral envelops of clearn and otherwise are somewhat analogous.

CONCLUSIONS

In this paper, we apply the method of signal decomposition and synthesis by wavelet transform and weighting coefficients for a specific level and cepstral analysis. This method is applied to the speech recognition. The speech recognition experiments by Japanese noisy digits are performed two kinds of colored noise are added to the original clear speech data for making noisy data. As a result, speech recognition rate is improved by this method. It was shown that the wavelet transform and weighting coefficients was one of the promising methodologies for filter.

REFERENCES

1 Stephane Mallat, "a wavelet tour of signal processing", Academic Press, 2nd Edit., (1998)
2 Paul S Addison, "The illustrated wavelet transform handbook", Institute of Physics Publishing Bristol and Philadelphoa, (2002)
3 D.E.Newland, "Random vibrations, spectral and wavelet analysis", Longman Scientific & Technical, 3rd Edit., (1993)
4 William H. Press, Saul A Teukolsky, Willam T. Vettrling, Brian P.Flannery, "Numerical Recipes in Fortran, The Art of Scientific Computing", Cambrige University Press, 2nd Edit., (1992)
5 Yves Nievergelt, "Wavelets Made Easy", Birkhauser, 2nd printing, (2001)
6 Rafael C. Gonzalez and Richard E. Woods, "Digital Image Processing", Addison-Wesley Publishing Company, (1993)
7 Y.Midorikawa, S.Hayano and Y.Saito, "A Thin Film Common Mode Noise Filter and Its Evaluation by Wavelet Analysis", IEEE Transaction on Magnetics, Vol.32, No.5, pp5001-5003, (1996)
8 Y.Midorikawa and M.Akita, "Improvement of the signal pattern recognition under the noisy environment by combining advanced spectral envelope modification and wavelet transform", Journal of Japan Society of Applied Electromagnetics and Mechanics Vol.12, Number 4, pp.310-315, (2004) (In Japanese)
9 S.Imai and Y.Abe, "Spectral Envelope Extraction by Improved Cepstral Method, Trans. IECE(A) Vol.62-A No.4, pp.217-223, (1979) (In Japanese)