Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

The Realtion between Speech Intelligibility and Aubibility by Voice Emphasis in the Subway

Taichi Ishigure (1), Yasukazu Kanamori (1)

(1) Aichi Prefectural Univesity, Aichi-gun, Japan

PACS: 43.66.Dc, 43.15.+s, 43.50.Cb

ABSTRACT

In the public transportation system such as subway, usually it is so noisy that passengers even can not hear the announcement like what the next stop and which door would be opened clearly. Obayashi, et. al. had studied the above mentioned case and found out that high frequency bands of the announcement would be cut off. To solve this problem, they proposed a method that extends the frequency of the voice. However, it was difficult to get good results for some consonants with the proposed method. Therefore, in this study, we newly propose a method to investigate and improve perceive consonants. First of all, we analyzed some announcements taken inside a subway, and compared the consonants by analysing power spectra of obtained announcenment. As a result we set up a hypothesis "Emphasizing the whole band of phoneme is the key point of effectiveness". Then, we carried out a hearing experiment in which the person under tested hear the obtained noisy voices and ask them what they heard. From the experiment, it is found that the recognition of consonants comparing with the pereceding study has been improved about 10% under keeping the voice quality clearly. We concluded that the propsed method in this study is effective for improving the consonant hearing in noisy environment.

1. INTRODUCTION

In daily life, many people use the public transportation such as bus, train, and taxi as their access tools. Therefore, the announcement that tells passengers about the next stop and the side of door opening is very important to hear, especially elderly and travler. Subway is one of the most popluar public transportation since it is always available even in bad weather and bad up-road traffic jam situation. However, it is so noisy inside of the subway that makes the announcement difficult to hear [1]. Ideally, the announcement should be clearly heard in all age groups.

From a preceding study [2], Obayashi et al found out that more than about 3.8 kHz of high frequency bands of the announcement were cut off. An example is shown in Figure 1, in which high frequency band of the announcement would be cut off at 3.8 kHz in the interval indicated by label "voice".

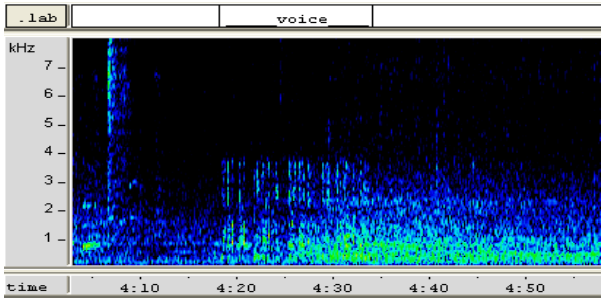

Figure 1. Spectrogram of voice inside the subway

Consequently, they set up a hypothesis "band limiting of the announcement causes the depression of the speech audibility". To verify their hypothesis, they had done a hearing experiment. The experiment carried out by letting each person under tested hear some noisy voices which are made up of different value of the band limiting (4 kHz, 6 kHz, 8 kHz) and ask them what they heard. As the experimental results, there were some improvements in voice recognition when the value of band limiting gets increased. However, this method is not effective to make some consonants to be recognized. Since there is at least one consonant in each subway station name in Japanese, it is very important to perceive conosonant.

In this study, we investigate a method to improve perceive level of voices, especially those consonants which was not able to get good result with preceding study with the voice emphasis.

2. SPECTRAL ANALYSIS OF VOICES AND NOISE

In this chapter, we will figure out the reason why the recognition of some consonants couldn't get good result by spectral analysis. Then, we will propose an optimal way of the voice emphasis from the results of the spectral analysis.

2.1 Power balance of the voices and noise

As a beginning, we have done the spectral analysis of voices and noise to see the power balance. We picked up the voices from ATR SPEECH DATABASE [2]. We used noise recroded in the subway. Both the voices and noise data are

same with the ones used in [1]. Setting and condition of the analysis is shown as Table 1.

Table 1. Setting and condition of the analysis

Application	Matlab
Cut off frequency	4kHz, 8kHz
Sampling frequency	48kHz

Results of the spectral analysis of the consonant /t/ are shown in Figure 2 and 3. In preceding study [1], they didn't get enough recognition of the consonant /t/. Figure 2 which has a 4 kHz cut-off frequency displays the power of the consonant being masked by the power of noise throughout whole frequency bands. Similarly, Figure 3 which has 8 kHz cut-off frequency displays most of the power of the consonant is still masked by the power of noise. Other consonatns that was not able to get enough reconigtion in preceding study show similar results with the consonant /t/ in the spectral analysis. In contrast, some consonants such as /s/ and /z/ which were imporved their recognition in preceding study generated different result from /t/. When the frequency band of the voice changed 4 kHz to 8 kHz, their power spectral got stronger than the power of noise between frequency band of 6 kHz and 8 kHz. Thus, we suggest that the audibility of the consonant was blemished because of the power of those consonants such as /t/ is masked by the power of noise throughout whole frequency bands.

.

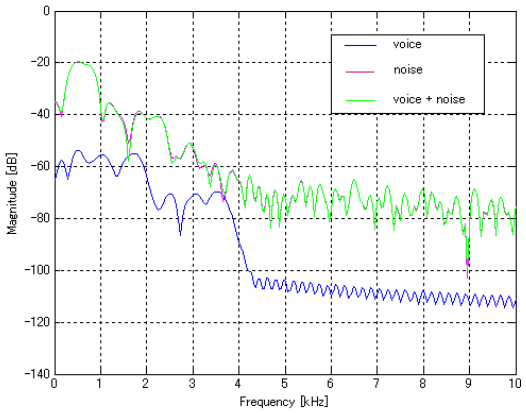

Figure 2. Power spectrum of the voice, noise and voice+noise when cut off frequency sets 4 kHz

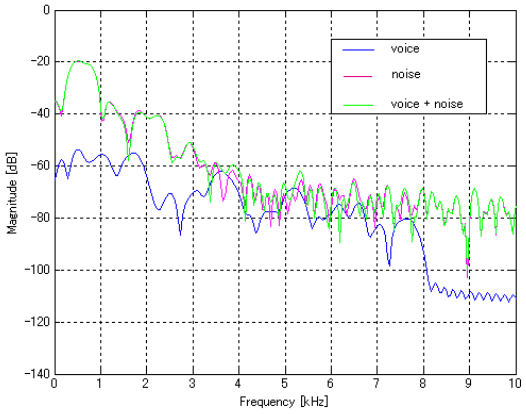

Figure 3. Power spectrum of the voice, noise and voice+noise when cut off frequency sets 8 kHz

2.2 Hypothesis under the spectral analysis

As the results of the spectral analysis, we set a hypothesis "Emphasizing the whole band of phoneme is the key point of effectiveness". When emphasize the amplitude of the voice, we have to consider about the distortion of the voice. Too much emphasis will cause distortion. Other way round, too weak emphasis will blemish the audibility. So we need to find the optimum degree of amplitude emphasis keeping the quality of phoneme.

3. A HEARING EXPERIMENT

From the hypothesis in previous chapter, we carry out a hearing experiment. We will show voice stimulus, experimental methodology and the assessment procedure.

3.1 Build process of voice specimens

As the voice stimulus, we made the voice that is added noise. Here, we will show the process of how to make the noise-include voice. Firtst, we devided the databese voice thorugh the phase dividing. Then, we picked up 6 phases that include the consonant /d/, /r/, /m/, /t/, /h/ at least. We did band limiting against each phases by using two different bands of low-pass filter (4 kHz, 8 kHz). Next, we adjust the amplitude of those phases by using SNR (-15 dB, -10 dB, -5 dB, 0 dB). Finally, we added the noise to those phases. We also use the voice that is not added noise but band limited (4 kHz) and emphasis the amplitude of voice to be SNR = 0 dB. As a result, we got a total of 108 voice stimulus for the hearing experiment.

3.2 Experimental methodology

First, people under test listened to the noise-included voices two times with headphone, and wrote down what the voice was heard on response sheets. Additionally, they need to rate the quality of the voice on a scale of 1 (unhear, not rudeness) to 5(too much hear, very rudeness). So the voice quality is proper when the value of loudness and rudeness are 3.0 and 2.0, respectively. Similar to actual subway environment, play each voice stimulus twice. The order of audio started from the band limited voice stimulus of 4 kHz to those of 8 kHz. A total of 8 people, 4 male and 4 female partipated this hearing experiment.

3.2 Summery procedure

By collecting the response of the people under test, we calculate the voice recognition level by using eqation (1). "Total" in EQ. (1) means "whole string". "Substitude error" means the "correct alphabet replace worng alphabet". "Dropout error" means "slip out the correct alphabet from the phase".

$$\text{Voice recognition}[\%] = \frac{\text{total} - \text{substitude error} - \text{dropout error}}{\text{total}} * 100 \quad (1)$$

4. EXPERIMENTAL RESULT

In this chapter, we show the results of the hearing experiment.

4.1 Trend of the voice recognition

The results of the recognition of syllabic, phoneme, vowel and consonant will be shown in next. The result of the band-limted voice stimulus of 4 kHz is shown in Figure 4. The horizontal axis of Figure 4 denotes the value of SNR. The vertical axis of Figure 4 denotes the recognition rate. The element "No-4k" in Figure 4 indicates the voice without noise. Figure 4 displays that the recognition rate gets more improved with the increase of SNR. The result of the band-limited voice stimulus of 8 kHz is shown in Figure 5. Figure 5 displays the similar finding to the result of 4 kHz. Figure 5 (8 kHz) demonstrate a better recongnition rate than, Figure 4 (4 kHz).

Next, we show the results of the recognition rate of each consonant in Figure 6 and Figure 7. Figure 6 is the result of band limiting 4 kHz. Figure 7 is the result of band limiting 8 kHz. Figure 6 shows that the recognition rate is improved when SNR gets bigger, especially voiced consonants /d/, /r/, and /m/. Figure 7 displays that most of the recognition rate excepting /h/ being improved to about 80% at the value of SNR being 0.

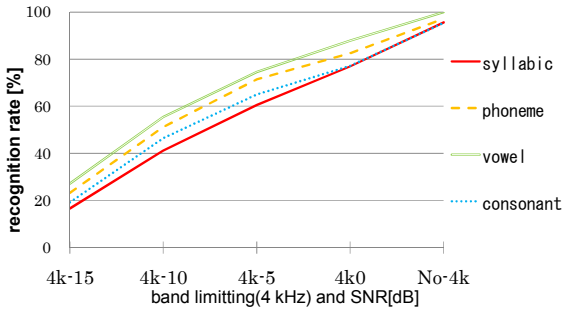

Figure 4. Result of the voice recognition in case of band limitting by 4 kHz

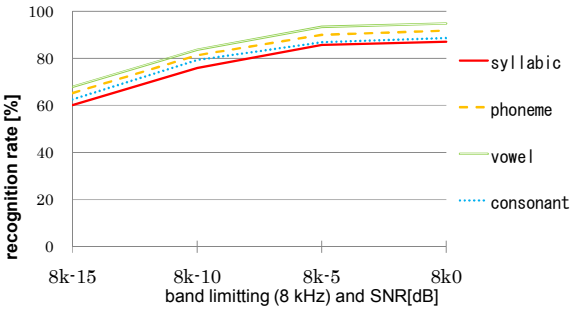

Figure 5. Result of the voice recgnition in case of band limiting by 8 kHz

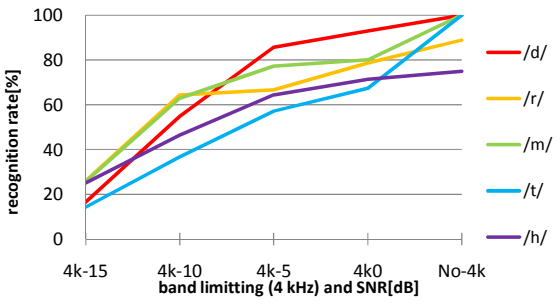

Figure 6. Recognition rate of the consonants in case of band limiting by 4 kHz

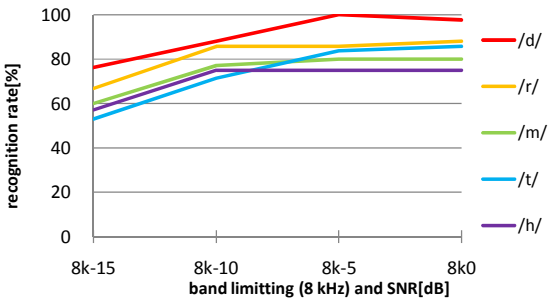

Figure 7. Recognition rate of the consonants in case of band limiting by 8 kHz

4.2 Results of the voice quality

Here, we show the result of voice quality (loudness and rudeness) value. First, show the case of band-limited by 4 kHz in Figure 8. Figure 8 shows the evaluated value of loudness is 3.7 when SNR equals to 0 dB. And the evaluated value of rudeness is 2.6 as Figure 9 shows. This means that the voice quality is not appropriate to hear. Other way round, when SNR is equal to -5dB, the evaluated value of loudness and rudeness are 3.1 and 2.1, respectively. So the voice quality is proper when SNR is -5dB in this case. Next, we show the result of the case of band limiting by 8 kHz in Figure 9. In this case, we understood that the voice quality is proper when SNR is equal to -10 dB.

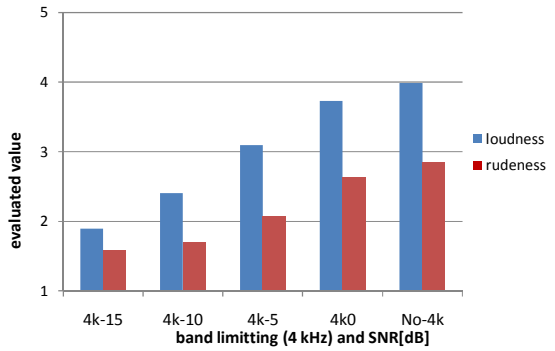

Figure 8. Evaluated value of loudness and rudeness with different SNR (4 kHz band limitting)

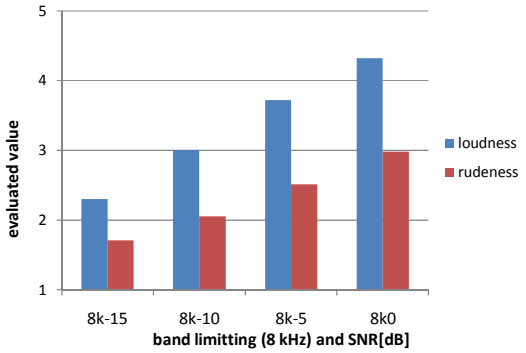

Figure 9. Evaluated value of loudness and rudeness with different SNR (8 kHz band limitting)

4.3 Comparison with preceding study

In order to comparing with results of preceding study [1], we use combined results of recognition rate and voice quality value. From the result of chapter 4.1 and 4.2, we could get the best suited set up of voice emphasis. The comparison between our results and those of preceding study is shown in Figure 10. From the Figure 10, we can see that the recognition of our results in the case of 8 kHz is higher than that of

preceding study, especially, the recognition rate of consonant /t/ is improved about 20%.

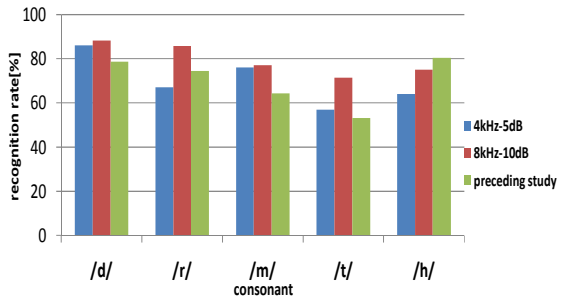

Figure 10. Compare the recognition rate of the consonants with preceding study

5. DISCUSSION

In this chapter, we discuss about the results from the experiment we have done.

The case of band limiting by 4 kHz has the similar set up with that of inside subway. We could get good recognition rate for consonants, /t/, /h/, /r/ in this case. However, an indication is obtained, i.e., if the amplitude of the annoucement set to the best value. It is difficult to obtain satisfied result. We found out that broaden the frequency range is very important for audibility.

From the experimental results, we understood that the voice quality would get ruder while SNR is increased. When we observe each response sheets from people under test, found that the voices of woman speaker tends to be more rudeness than the voice of man speaker. The voice of woman speaker is used as the announcement inside of the subway. As known from equal-loudness curve, we can hear high frequency sound better than low frequency sound [4]. Therefore, it is highly possible that the announcement will be rudeness when the amplitude of the announcement was emphasised, which is our case. Through other detailed experiment, it would be get the best suited set up of the degree of amplitude emphasis.

The recognition rate of the consonant /h/ from proposed method is lower than the recognition rate from preceding study. In Figure 6, the recognition rate of /h/ is the lowest when the element is “No-4k”. “No-4k” is the voice that is not added noise but band limiting (4 kHz) and emphasis (SNR = 0 dB). When we check each resoponse sheets of people under test, they confused the consonant /h/ with the consonant /k/. It would appear that too emphasis /h/ like set up SNR equal to 0dB cause perceptual error. Therefore, the other case “8k-0” get scarce recognition rate. We also check each response sheet of subjects when we set up SNR equal to -15dB, -10dB, -5dB. Then, we obtained that the most of subjects is not so much perceptual eroor as unhear the consonant /h/. The consonant /h/ has 2 features. It has the formant that is the resonance of vocal tract. But it does not have the antiresonance [5]. Therefore, it would appear that a vowel after a consonant will be very important for audibility. In order to improve the audibility of the consonant /h/, it could be possible that emphasis the local part of frequency band of /h/ same as the formant frequency of the vowel after the consonant /h/. At that time, we need to set up the amplitude of the voice to resist rudeness.

6. CONCLUSION

In this study, we investigated the trend of speech audibility when emphasis the whole amplitude of phoneme by changing SNR. As a result, we could get the best suited set up of the amplitude emphasis level keeping the band limetted voice (4 kHz, 8 kHz) quality. Especially, the result of the case of 8 kHz, we could get higher recognition rate than the rate of preceding study. It is found that the hypothesis “Emphasizing the whole band of phoneme is the key point of effectiveness” is correct.

We have shown the improved audibility of the announcement inside of the subway by broaden the frequency range. However, further effort is needed to complete our method because some consonants make very little improvement. Furthermore, there is at least one consonant in each subway station name in Japan. It means that the audibility of the consonant is very important for recognition of subway station name. So we propose the way that is emphasis the amplitude along with broaden the range of frequencies. We conclude from the experiment described above that the audibility of the consonant can be improved keeping their speech quality by using the way we proposed even in noisy environment.

7. REFERENCES

1 Ministry of Land (2006) “A summary of public comment about architecture of the car and basis of the equipment or the passenger plant for smooth traveling” (http://www.mlit.go.jp/pubcom/06/kekka/pubcomk90/01.pdf)

2 Y. Obayashi and Y. Kanamori, “The audibility of band-limited announcement in the subway inside of car” *Acoustical Society of Japan*. **56**, 1-3-3 (2008)

3 ATR-Promotions (1992), “voice database - voice and parallel translation text data” (http:www.atrp.com/sdb.html)

4 M. Kasuga, T. Hunada, S. Hayashi and K. Takeda, *Speech and Sound Signal Processing for Human Interface* (corona publishing, Tokyo, 2001) p. 162

5 K. Shikano, S. Nakamura and S. Ise, *Speech Digital Signal Processing* (Shokodo, Tokyo, 1997) p. 29