Analysis of Speech Depending on Age and Speaking Style by Using Acoustic Features

Shunsuke Kawahara^{a,*}, Yoshihisa Nakatoh^a

^aKyushu Institute of Technology

1-1, Sensui-cho, Tobata-ku, Kitakyusyu-shi, Fukuoka, 804-0015, Japan

*nakatoh@ecs.kyutech.ac.jp

Abstract

Currently, aging is progressing and it is expected that opportunities of elderly people to communicate will increase. Also, the ease of listening announcement depends on the age of the speaker and the speaking style.

In this paper, we analyze and compare the acoustic features of speech uttered by the adult (adult speech) and speech uttered by the elderly (elderly speech). And, we analyze and compare the acoustic features of clear speech uttered (clear speech) and speech uttered as usual (normal speech).

As a result, it is found that the adult speech has larger intensity and spectral flux in the specific band than the elderly speech. Also, it is found that clear speech has larger volume and spectral flux than normal speech.

Keywords: announcement, acoustic features, elderly people, speaking style

1. Introduction

Currently, aging is progressing and it is expected that opportunities of elderly people to communicate will increase. Also, the ease of listening announcement depends on the age of the speaker and the speaking style.

In this paper, we analyze and compare the acoustic features of speech uttered by the adult (adult speech) and speech uttered by the elderly (elderly speech). And, we analyze and compare the acoustic features of clearly speech uttered (clear speech) and speech uttered as usual (normal speech). Also, we investigate the acoustic features affecting listening from the comparison results.

2. Acoustic features

The acoustic features analyze in this study are volume feature and tone color features. The volume feature is the intensity, the tone color features are spectral contrast, spectral centroid, spectral roll off, spectral flux [1]. In this chapter, we explain that the features extraction method and the details of the parameters used in this study. The acoustic features are calculated from the short-time power spectrum obtained by short-time Fourier transform (STFT). The STFT conditions are the sampling frequency is 16 kHz, the frame width is 256 ms, the frame shift is 8 ms, and the window function is Hanning window. The intensity, the spectral contrast and the spectral flux are divided into 8 ($n = 8$) using an octave scale band to divide the power spectrum into the subbands to analyze in detail. Also, we use the following formula for the division.

$$\left[0, \frac{fs}{2^n}\right], \left[\frac{fs}{2^n}, \frac{fs}{2^{n-1}}\right], \dots, \left[\frac{fs}{2^2}, \frac{fs}{2^1}\right] \quad (1)$$

Therefore, the subbands are 0-62.5 Hz, 62.5-125 Hz, 125-250 Hz, 250-500 Hz, 500-1000 Hz, 1-2 kHz, 2-4 kHz, 4-8 kHz.

2.1 Intensity

Intensity is the volume feature. $I(t)$ is the volume of the entire frame, and $D_j(t)$ is the volume ratio of each subband that divided the frame.

$$I(t) = \sum_{f=0}^{fs/2} X(t, f) \quad (2)$$

$$D_j(t) = \frac{1}{I(t)} \sum_{f=L_j}^{H_j} X(t, f) \quad (3)$$

Where $X(t, f)$ is the power spectrum, t is the frame number, f is the frequency, j is the subband number, L_j is the lower limit frequency of the subband and H_j is the higher limit frequency of the subband.

2.2 Spectral contrast

Spectral contrast C_j is the feature of voice clarity. Spectral contrast is calculated by the ratio of the subband peak P_j to the subband valley V_j . We use X_j sorted in descending order by the size of the power spectrum in the j th subband to calculate subband peak and subband valley. Also, they are calculated by the following formulas.

$$P_j = \log \frac{1}{\alpha N} \sum_{k=1}^{\alpha N} X'_{j,k} \quad (4)$$

$$V_j = \log \frac{1}{\alpha N} \sum_{k=1}^{\alpha N} X'_{j,N-k+1} \quad (5)$$

$$C_j = P_j - V_j \quad (6)$$

Where N is the number of discrete frequency points in the j th subband and k is the element number in the j th subband. Also, α is the ratio of the subband peak and the subband valley contained in the subband, and $\alpha = 0.2$.

2.3 Spectral centroid

Spectral centroid $f_c(t)$ is the feature of sound brightness. The spectral centroid is calculated by the following formula.

$$f_c(t) = \frac{\sum_{f=0}^{fs/2} X(t, f) \cdot f}{\sum_{f=0}^{fs/2} X(t, f)} \quad (7)$$

2.4 Spectral roll off

Spectral roll off f_r is the feature of shape of the

spectrum. Spectral roll off is the frequency occupying 95% of the entire band in the spectral distribution, and it is calculated by the following formula.

$$\sum_{f=0}^{f_r} X(t, f) = 0.95 \times \sum_{f=0}^{fs/2} X(t, f) \quad (8)$$

2.5 Spectral flux

Spectral flux is the feature of sound fluctuation. The spectral flux F is the entire spectral flux, and F_j is the spectral flux of each subband. The spectral fluxes are calculated by the following formulas.

$$F = \sqrt{\sum_{f=0}^{fs/2} (X(t, f) - X(t-1, f))^2} \quad (9)$$

$$F_j = \sqrt{\sum_{f=L_j}^{H_j} (X(t, f) - X(t-1, f))^2} \quad (10)$$

Where j is the subband number, L_j is the lower limit frequency of the subband, and H_j is the higher limit frequency of the subband.

3. Comparison of acoustic features

This chapter explains the recording of adult speech and elderly speech, comparison methods, comparison results and consideration.

3.1 Recording

We record clear speech and normal speech of 5 adults (21-22 years old) and 5 elderlies (70-77 years old) in the soundproof room.

We teach the subjects two styles of clear speech and normal speech. And, we record 50 words of 4 mora words [2] with familiarity 1.0 - 2.5.

3.2 Comparison method

The acoustic feature of the word sound is calculated for each frame. We calculate the frame average of each feature by dividing the sum of feature quantities with the number of frames. Next, to each speaker and speaking style, the average of the each feature quantities are calculated from all word sounds. Finally, the quantities of each feature are

normalized so that the maximum value becomes 1, and gives a comparison.

3.3 Comparison results

As a comparison between adult speech and elderly speech, Fig.1 to Fig.4 show average of the acoustic features of adult speech and elderly speech. Also, the numbers on the horizontal axis of these figures indicate subband numbers. Table.1 shows the subband numbers.

As a result, from Fig.1, it is found that clear speech of both adults and elderlies has larger volume than normal speech, but it is smaller in the low frequency range (62.5-125Hz). Also, it is found that adult speech has larger intensity in the low frequency range (62.5-125Hz) and the high frequency range (4-8kHz) than elderly speech. On the other hand, we find that elderly speech has larger intensity in the middle frequency range (250-500Hz) than adult speech.

Fig.2 shows that adult speech has larger spectral centroid and spectral roll off than elderly speech.

Fig.3 shows that clear speech of both adults and elderlies has larger spectral flux than normal speech, but it is smaller in the low frequency range (62.5-125Hz). Also, it is found that adult speech has larger spectral flux in the low frequency range (62.5-125Hz) and the high frequency range (4-8kHz) than elderly speech. Further, it is found that clear speech of adults has larger spectral flux in the 1-8 kHz band than clear speech of elderlies.

Fig.4 shows that clear speech of both adults and elderlies has smaller spectral contrast in the low frequency range (62.5 to 125 Hz) than normal speech.

Fig.1 Intensity of adult speech and elderly speech

Fig.2 Spectral centroid and spectral roll off of adult speech and elderly speech

Fig.3 Spectral flux of adult speech and elderly speech

Fig.4 Spectral contrast of adult speech and elderly speech

Table1. Subband numbers

Subband number	Frequency band
1	0-62.5Hz
2	62.5-125Hz
3	125-250Hz
4	250-500Hz
5	500-1000Hz
6	1-2kHz
7	2-4kHz
8	4-8kHz

3.4 Consideration

From Fig.1, it is found that clear speech of both adults and elderlies has larger volume than normal speech, but it is smaller in the low frequency range (62.5-125Hz). From the result, it is thought that speaking by being aware of ease of listening is the voice becomes bigger, and improves the indistinct voice of low frequency. Also, it is found that adult speech has larger intensity in the high frequency range (4-8kHz) than elderly speech. From the results, it is thought that adults utter consonants stronger than elderlies because frequency range of consonants is mainly in the high frequency range. On the other hand, it is found that elderly speech has larger intensity in the middle frequency range (250-500Hz) than adult speech. From the results, it is thought that elderlies utter vowels stronger than adults because frequency range of vowels is mainly in the middle frequency range.

From Fig.2, it is found that adult speech has larger spectral centroid and spectral roll off than elderly speech. From the results, it is found that adult speech is brighter than elderly speech, and it is thought that it is caused by the difference in the intensity of the high frequency range.

From Fig.3, it is found that clear speech of both adults and elderlies has larger spectral flux than normal speech, but it is smaller in the low frequency range (62.5-125Hz). From the results, spectral flux is increasing / decreasing trend similar to intensity so that we think that they are deeply related features. Also, from the previous study [3], it is

known that clear speech has larger dynamic feature than normal speech. The spectral flux used in this paper is also dynamic feature so that it is thought that results consistent with the previous study are obtained. However, the spectral flux in the low frequency range of clear speech becomes smaller so that it is found that dynamic feature of clear speech hasn't larger in the whole band. Further, it is found that clear speech of adults has larger spectral flux in the 1-8 kHz band than clear speech of elderlies. From the results, it is thought that dynamic feature in the high frequency range is particularly important.

From Fig.4, it is found that clear speech of both adults and elderlies has smaller spectral contrast in the low frequency range (62.5 to 125 Hz) than normal speech. The spectral contrast is the feature of voice clarity so that it is thought that high clarity of low frequency range may be cause sound difficult to listen.

Conclusion

We analyze and compare the acoustic features of adult speech and elderly speech. And, we analyze and compare the acoustic features of clear speech and speech normal speech. As a result, it is found that the adult speech has larger intensity and spectral flux in the low frequency range (62.5-125Hz) and the high frequency range (4-8kHz) than the elderly speech. Also, it is found that clear speech has larger volume and spectral flux than normal speech, but clear speech has smaller both of the acoustic features in the low frequency range (62.5-125Hz) than normal speech.

In the future, we plan to investigate the relevance between the numerical results obtained and the listening impression. Furthermore, we plan to examine the speech processing method for improving the clarity of speech by referring the analysis results obtained in this paper.

References

- [1] Y. Hiraga et al., "Method for Calculating the Subjective-based Music Similarity Measure", Information Processing Society Research Report, pp1-6, 2009.7
- [2] S.Sakamoto et al., "New lists for word intelligibility test based on word familiarity and phonetic balance", Journal of Japan Acoustical Society, pp842-849, 1998.12
- [3] Y.Sibata et al., "Improvement method of intelligibility of speech analysis and synthesis method focusing on the dynamic feature Guide to the Technical Report and Template", IEICE technical report. Speech, pp842-849, 1998.12