Investigation on variation of the initial phonemes of Japanese words due to aging

Yoshiaki Yuda ^{a,*}, Ryo Fukuhara ^a, Mitsunori Mizumachi ^a, Yoshihisa Nakatoh ^a

Kyushu Institute of Technology

^a 1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka 804-8550, Japan

* Yoshiaki Yuda: k349437y@tobata.isc.kyutech.ac.jp

Abstract

Speech recognition system is infiltrated widely in our daily life. However, speech recognition rate for elderly speech are lower than that for non-elderly speech. Therefore, in this paper, we analyzed elderly speech so that recognition rate for elderly people becomes higher.

In previous work, we undertook to find out whether major subjective characteristics for elderly speech are “rough”, “slow speaking”, and “non-briskness”. Furthermore, each subjective characteristics have an acoustic characteristics corresponded. In those experiments, on medial phoneme, elderly speech was analyzed. However, a medial phoneme has an influence on coarticulation. Therefore, in this paper, on initial phoneme which has only a slight influence on coarticulation, we investigate relationship between subjective characteristics and acoustic characteristics for elderly people.

In the result, aging influence on initial phoneme is shown. For instance, duration of initial phonemes is getting longer at the steady-state phonemes. Power spectrum lifts in high frequency band of a voiced sound. And Cepstrum distance decreased.

Keywords: Speech analysis, Elderly speech, Rough, Slow speaking, Non-briskness

1. Introduction

Although speech recognition systems have infiltrated many aspect of our daily life, recognition rate for elder people isn't larger than those for non-elderly. Therefore, it is important to investigate the characteristics of elderly speech to build elderly-friendly speech interfaces. In previous work, Characteristics of word utterances by elderly people, who

are defined here as people over 60 years old, was analyzed. Major subjective characteristics for elderly speech express “rough”, “slow speaking”, and “non-briskness” [1].

In this study, we analyzed the initial phonemes of Japanese words. Through analyzing the initial phonemes, we can investigate the influence of aging at Japanese phonemes except the influence of phoneme bond currently taken into consideration by the previous work. Then, we analyzed what kind of acoustical dynamic features expresses. And, we investigate what kind of relevance is between the characteristics of elderly speech and the acoustical feature of initial sound matter.

2. Subjective characteristics of elderly speech

We conducted a listening test to determine the degree of subjective characteristics for elderly speech that include “rough”, “slow speaking”, and “non-briskness”[1]. The subjects were 10 male and 10 female students, and they were asked to judge the degree of subjective characteristics on a five-point scale from 1 to 5, where a higher score indicated a less characteristic voice. The subjects listened to 50 connected words, which were prepared from a phonetically balanced 543-word database, spoken individually by 36 elderly male speakers. Each speaker was labeled with the degree of characteristics based on the five-point scale. Each figure of the degree of characteristics for elderly speech in "slow voice", a "rough voice", and "non-briskness" is carried as Fig. 1. . A vertical axis expresses the evaluation degree in each feeling impression of characteristics for elderly speech, and a horizontal axis rearranges each evaluation speaker's value in order of an evaluation degree.

In order to compare with elderly speaker, six 20 years-old adult men speaker were prepared, and it used for analysis.

3. Analysis of slow voice by elderly speakers

We analyze the slow voice, which is one of the important characteristics of elderly speech, objectively.

3.1 Duration of initial phonemes

“Slow speaking” voices are perceived by talking slowly in order to compensate ambiguous pronunciation. So we calculated duration of initial phonemes gropes and compared slow voiced speakers with non-slow voiced speakers and 20’s each other. Duration of initial phonemes is calculated as average value of all the speakers in each group, i.e. slow-voiced speakers and so on. We calculate duration of initial phonemes by using the phoneme enclosure information on the label data given to the data of the phonetically balanced 543-word database.

The duration of initial phonemes is calculated per frame and 1 frame is 10 milliseconds.

3.2 Relationship between subjective characteristics of slow voice and Duration of initial phonemes

In the result of experience, as the speakers who speech slowly, duration of initial phonemes is getting longer at the steady-state phonemes such as vowel, nasal, and buzz of voiced plosive and voiced plosive. But then, duration of initial phonemes don’t prolong at the non-steady-state phoneme which is unvoiced plosive. Duration of initial phoneme calculated in each phoneme groups is shown in Fig. 2. These show that the influence of fall of muscular power by aging is seen by articulator.

4. Analysis of rough voice by elderly speakers

We analyze the rough voice, which is one of the important characteristics of elderly speech, objectively.

4.1 Calculates Average spectrum of Initial phoneme

“Rough” voices are perceived by overhearing the air which leaked when vocal cords damaged by aging.

So we calculated average spectrum of initial phoneme and compared rough voiced speakers with non-rough voiced speakers and 20’s each other. Average spectrum of initial phoneme is calculated as average value of all the

(a) Preference Score of slow voice.

(b) Preference Score of rough voice.

(c) Preference Score of brisk voice.

Fig. 1. Preference Scores of characteristics of elderly speech.

speakers in each group, i.e. slow-voiced speakers and so on. We calculate average spectrum of initial phoneme by using the phoneme epoch information on the label data given to the data of the phonetically balanced 543-word database.

4.2 Relationship between subjective characteristics of rough voice and Average spectrum of Initial phoneme

In the result of experience, as the speakers those who speak by rough voice, energy of power spectrum lifts in higher frequencies in the initial phoneme produced with vibration of the vocal cords, such as vowels and buzz of voiced plosive and voiced plosive. But then, there doesn't verify that energy of power spectrum lifts in higher frequencies at in the initial phoneme produced without vibration of the vocal cords, such as unvoiced plosive and unvoiced fricative, and so on. Figure 3 shows average spectrum of initial phoneme at vowel /a/ and unvoiced fricative /h/ for each speaker group. These results show that the influence of air which leaked from vocal cords by damage to the vocal cords by aging is seen by articulator.

5. Analysis of non-brisk voice uttered by elderly speakers

We analyze the non-brisk voice, which is one of the important characteristics of elderly speech, objectively.

5.1 Cepstrum distance increase on initial phoneme

In an earlier experiment, on medial phoneme, we undertook to find out whether spectrum shift of non-brisk voice is small on transition position, then, adjacent phonemes join is obscure. In this paper, Cepstrum distance increase on initial phoneme is calculated to exclude coarticulation influence. It is defined by Eq(1).

$$dif_{spectrum} = \sqrt{\sum_{n=1}^{30} (C_n^{(t_1)} - C_n^{(t_2)})^2} \quad (1)$$

where C represents cepstrum coefficient by FFT, t1 and t2 are frame indexes, and the order of cepstrum analysis is 30.

Initial phonemes consist of silent zone-vowel or silent zone-consonant. Therefore, the frame indexes t1 is determined to be a frame in silent zone, and the frame indexes t2 is determined to be the most distinctive positions of the target phoneme. As the feature point of each phoneme, we choose the center of each phoneme for

Fig. 2. Duration of Initial phonemes.

(a) Average spectrum of Initial phoneme /a/.

(b) Average spectrum of Initial phoneme /h/.

Fig. 3. Average spectrum of Initial phoneme.

stationary phonemes such as a vowel and a fricative; it depends on the kind of phoneme for non-stationary phonemes. For example, it should be set as the plosive

Fig. 4. Cepstrum distance increase on initial phoneme position for plosives.

5.2 Relationship between acoustic characteristics of non-brisk voice and Cepstrum distance increase on initial phoneme

Cepstrum distance increase on initial phoneme calculated in each phoneme groups is shown in Fig. 4. From Fig. 4., on all phonemes except voiced fricatives, Cepstrum distance increase on initial phoneme of non-elderly voice is larger than that of elderly voice. Furthermore, Cepstrum distance increase initial phoneme of brisk voice is larger than that of non-brisk voice.

Cepstrum distance increase on initial phoneme is smaller with growing age, in particular for non-brisk speaker. This is because it is difficult to articulate appropriately for the elderly speaker. In voiced fricatives, because the number of data to use is too few, Cepstrum distance increase on initial phoneme of non-brisk speaker is smaller than that of non-elderly speaker.

6. Conclusions

In this paper, on initial phoneme, duration of initial phonemes, power spectrum, and Cepstrum distance increase (which corresponds respectively to slow utterance voice, hoarse voice, and non-brisk voice) is analyzed to investigate aging influence on initial phoneme. In the result, aging influence on initial phoneme, such as duration of initial phonemes increased, power spectrum increased on high frequency band of a voiced sound, and Cepstrum distance decreased, is shown. This result to analyze on initial phoneme is similar to that in medial voice.

In future works, we will investigate quantification of

preference score of 20's voice and analysis of acoustic feature because 20's voice might be rough voice for having a smoke.

Acknowledgment

This work was partially supported by Grant-in-Aid for Science Research (No.19560387) from the Japan Society for the Promotion of Science.

References

- (1) T.Miyazaki et al., "Study on auditory impressions of elderly speech," Proc. ASJ meeting, pp. 427-428, 2008.
- (2) D.Harada et al., "Analysis of ill-articulated speech by elderly speakers based on temporal transition of amplitude spectrum" Proc.NCSP,