Comparison of nonverbal feature of free conversation speech between elderly and young individuals

Yumi Wakita ^{a,*}, Masaya Yamazoe ^a, Yoshihisa Nakatoh ^b

^aOsaka Institute of Technology, Osaka, Japan

^bKyushu Institute of Technology, Kitakyushu, Japan

*Yumi.wakita@oit.ac.jp

Abstract

This study explores a system that estimates the degree of decline for elderly persons using nonverbal information from daily conversations. Factors useful in estimating the degree of decline and methods of estimation need to be determined.

First, we confirmed whether we can estimate the speaker's age and "elderly likeness" from the conversation data. We made several processed conversation sounds: ("only included Fillers," "only included Laughter utterances," and "only included Intonation information") from the original conversation database. By listening to these sounds, we confirmed that it is possible to estimate whether the speaker is young or elderly.

Next, we analyzed the nonverbal features of conversation sounds of young and elderly individuals. In this paper, we selected the F0 and power value of "laughter utterance" and "speech utterance," and compared the difference between young and elderly individuals.

By comparing the results, we discussed the possibility of estimating a degree of decline.

Keywords: degree of decline, conversation sounds, laughter utterance, fundamental frequency.

1. Introduction

Although the number of accidental deaths has decreased, the traffic death accident rate for elderly people increases every year. The primary cause is "driving inappropriately". Although there is a decline in ability, that

they overestimate their ability to that when they were young, without awareness of decline. Many proposals for dementia prevention and health maintenance measures have been proposed; however, without a recognition of this problem, prevention measures are unclear. A serious accident leading to death can occur suddenly. It is very important for third parties to perceive their decline and to be aware of it.

We are in the process of developing a system that can estimate the degree of decline for an elderly person from his or her daily conversation, and inform one's degree of decline when the degree will decrease rapidly. To create this system, it is necessary to find factors that are effective in estimating the degree of decline.

Several papers illustrated that the acoustic features of the human voice are effective in detecting diseases specific to the elderly. Kato et al.⁽¹⁾ reported that the prosodic feature is effective in estimating cognitive impairment in the elderly. Taler et al.⁽²⁾ reported on the relation between Alzheimer's disease and disorders as found in the prosody. Some characteristics of speech among the elderly were reported. Formant frequency shifts in the speech of the elderly were found by Tanaka et al.⁽³⁾. The purpose of estimating the degree of decline is to determine it before it becomes a serious disease. Most analyses did not use free conversations. To notice this impairment in daily life, we should analyze daily conversations.

We already confirmed that F0 behaviors in free dyadic conversations are effective in estimating a conversational atmosphere. The standard deviation value of F0 (SD-F0) for each utterance in a smooth conversation is greater than that of a non-smooth conversation⁽⁴⁾⁽⁵⁾. It is necessary to confirm whether F0 information is useful in estimating not only the conversational atmosphere but also in estimating the degree

of age decline.

In this paper, in section 2, we determine whether we can estimate the speaker's aging by listening to conversation sounds and determine the acoustic features that are useful for estimation. In section 3, we analyze the nonverbal features of conversation sounds and discuss the differences between young and elderly individuals.

2. Possibility of age estimation using nonverbal feature of conversation sound

2.1 Conversation database

We recorded 18 sets of 3-min free dyadic conversations (between two individuals). Fig. 1 shows the conversation recording location. We used two microphones and a video camera for recording. The recording conditions are listed in Table 1. The participating speakers had met each other before, but we used pairs who had never spoken to one another.

Fig. 1. Conversation recording location

Table 1. Conversation database

Speakers	Young (6 males, 2 females) Elderly (6 males, 4 females)
Ages	Young individuals: 22 years old Elderly individuals: 65–82 years old
Conversations	18 pairs (10 elderly pairs, 8 young pairs)
Conversation periods	3 in/conversation
Conversation condition	Free dyadic conversation

2.2 Effectiveness of nonverbal information of speech for age estimation

We indicated the original conversation sound source to five persons and asked how old the speakers were.

In addition, we made three kinds of processed sound source from original conversation. The processing methods were as follows:

- Filler utterances (e.g., “ah” and “un”) selected from the original conversation database.
- Laughter utterances selected from the original conversation database.
- Speech data processed through a low-pass filter with a cut-off frequency of 400 Hz.

The data (A) and (B) include the same acoustic characteristics as the original data, but do not include language information. For data (C), both the language information and the acoustic parts that are important to recognize language contents are removed from the original. Only the intonation is retained. We indicated the data (A)–(C) to five persons and asked them to estimate the speaker's age.

2.2.1 Discriminating between elder sounds and the elderly

We prepared two cards with the age categories “20s” and “over 60.” We asked five persons to estimate the age of the speakers just after listening to a conversation, and to select the appropriate card. Twelve speakers were extracted from the database in Fig. 1. Six speakers were young, and six speakers were elderly (two in their 60s, two in their 70s, and two in their 80s). Each length of conversation was 10 s. Table 2 lists the correct answer rates for each age.

Table 2. Correct answer rates for each age

	Original	(A)	(B)	(C)
Young (6 males)	1	0.975	0.967	0.867
Elderly (6 males)	0.983	0.95	0.965	0.9

- The correct answer rates of any database were high scores. The rates were over 86.7%.
- The correct answer rates decreased when using data (C).
- The rate for elderly speakers were lower than youngers, but only when using (C), the elderly's were higher.

The results suggest the following:

- Discriminating between the elderly and the young is possible without language information.
- Only intonation information is effective in estimating the age. However, the estimation performance decreases without an acoustic part over 400 Hz.

2.2.2 Age estimation for elderly individuals

We prepared three cards with the age categories “60s,” “70s,” and “80s,” and asked the other five persons to select the appropriate card after listening to the elderly conversation database. The length of each conversation was 10 s. We indicated two different conversation parts for each condition. Table 3 lists the correct answer rates.

Table 3. Correct answer rates for each age

	Original	(A)	(B)	(C)
60s	0.75	0.75	0.70	0.65
70s	0.40	0.35	0.30	0.35
80s	0.50	0.50	0.75	0.40

- The correct rates were not high. Ages could not be estimated, especially for the speakers in their 70s.
- For the speakers in their 80s, when using laughter database (B), the rates increased.
- When using only the intonation information database (C), the correct rates decreased. The tendency was the same as the result in chapter 2.2.1.

The results indicate it is difficult to estimate the ages of elderly individuals using conversation sound information. However, the incorrect tendencies are almost the same for all persons.

Fig. 2 shows a histogram of answers by five persons for the original database. Speaker1 and Speaker2 are in their 60’s ages, Speaker3 and Speaker4 are in their 70’s ages, and Speaker5 and Speaker6 are in their 80’s ages. The correct rates are dependent of speakers and for some speakers, the rates are low. However, for speaker3, who was in his 70s, almost all persons answered that he was in his 80s. For speaker1, who was in his 60s, almost all persons’ answers were correct. The tendency of the answers is similar for each person. For the original data, and for (A) (B) (C), the tendencies are the same.

The results indicate that we can determine the degree of decline using the same factors in the database. The factors that are useful in estimating the degree of decline are included in the nonverbal information of conversation sounds.

Fig. 2. Number of answers by five persons for original database

3. Comparison of nonverbal information between young and elderly

3.1 Laughter utterances

From the results of the experiments in chapter 2.2.1, the laughter utterances are more effective than typical speech for age estimation. We analyzed the laughter utterances by comparing young and elder individuals.

Table 4. Ratio of laughter utterances to all utterances

	Ave. of all speakers	Standard Deviation of all speakers
Young	0.191	0.054
Elderly	0.187	0.046

Table 5. Ratio of each type for all laughter utterances

	Pleasantness	Pleasantness with speech	Sociability	Sociability with speech
Young	0.388	0.102	0.45	0.06
Elderly	0.273	0.0	0.49	0.237

Table 6. Ratio of utterances where F0 could be extracted relative to all laughter utterances

	Ave. of all speakers	Standard Deviation of all speakers
Young	0.95	0.05
Elderly	0.68	0.15

Table 4 shows the ratio of laughter utterances to all utterances. Table 5 shows the ratio of each type for all laughter utterances when we classified the laughter utterances as four types: “pleasantness”, “sociability”, “pleasantness with speech”, or “sociability with speech”⁽⁶⁾. Both the ratio of laughter utterances and the ratio of each type are similar when comparing elderly with young individuals.

Table 6 shows the ratio of utterances where F0 could be extracted relative to all laughter utterances. Almost all young laughter utterances, including the parts of voicing and F0, can be extracted. However, for elderly laughter utterances, 32% of laughter utterances are unvoiced, and F0 cannot be extracted. This indicates that laughter utterances change from voiced to unvoiced as people age.

Fig. 3. Distribution of Ave-F0 and SD-F0 of laughter utterances

Fig. 3 shows the distribution of the Ave-F0 and SD-F0 values of laughter utterances for all speakers. The results show the following:

- The area for the elderly is smaller than that for young individuals. The utterances of the elderly are plotted almost in the young individual’s area.
- When both the Ave-F0 and the SD-F0 values of a laughter utterance are large, this indicates that the utterance is from a young individual.

3.2 Speech utterances

From the experimental results described in chapter 2.2.1, we can estimate age using only the acoustic characteristics without language information laughter

utterances. We analyzed F0 of the speech utterances by comparing young and elder individuals.

Fig. 4. Distribution of Ave-F0 and SD-F0 of speech utterances

Fig. 4 shows the distribution of the Ave-F0 and SD-F0 values of speech utterances for all speakers. The results show the following:

- The SD-F0 of elderly persons is slightly higher than those of younger persons. The difference is significant when using a t-test, where the confidence level is 95%.
- The area for the elderly is smaller than that for young individuals. This tendency is similar to that of laughter utterances.

In Fig. 2, the speech data of Speaker1 is regarded as sound from a person in his 60s by almost all listeners. On the other hand, the speech data of Speaker3 are regarded as sounds from a person in his 80s by almost all listeners. Fig. 4 are comparing

Fig. 5 shows the distribution of SD-F0 and the standard deviation of power (SD-Power) for Speaker1 and Speaker3.

Fig. 5. Distribution of SD-F0 and SD-Power of speech utterances by Speaker1 and Speaker3

Fig. 5 indicates that the SD-F0 and SD-power of Speaker 1's utterances are slightly higher than those of Speaker 3. The differences are not significant when using a t-test. However, some utterances of speaker1 indicated an extremely high F0 or large power values. We have the impression that individuals are young from these extreme values.

4. Conclusion

To realize a system that can estimate the degree of decline from daily conversations, we analyzed the acoustic features of conversation sounds and described the effectiveness for estimating the degree of decline for elder individuals.

We confirmed that when we estimate the age of speakers, the acoustic characteristics of laughter utterances or the intonation information is useful. By analyzing conversation sounds, we confirmed some different points between utterances of the elderly and the young.

The laughter utterances become unvoiced sound as getting older. When F0 can be extracted from laughter utterances, the standard deviation values of F0 of elderly laughter utterances are smaller than those of young persons.

The speech utterances have a tendency that the dynamic range of SD-F0 for elderly individuals tends to be narrower when compared with that of young individuals. These results indicate that conversation atmosphere estimation for elderly individuals using F0 characteristics is more difficult when compared with estimating for young individuals.

Comparing the SD-F0 and SD-power of the utterances by Speaker1 who is regarded as in his 60's age with

utterances by Speaker3 who is regarded over 80 years, both of the SD-F0 and the SD-power of Speaker3 are slight lower than Speaker1's. The results indicate when we estimate an age by voice, the width of the standard deviation of F0 and power values of each utterance have possibilities as effective factors.

We confirmed the relationship between aging and the conversation sound characteristics in this paper. However, in the future, it is necessary to confirm the effectiveness of these acoustic characteristics not only by using an aging estimation but also by examining the degree of decline. Also the size of the database is not large. In the future, we aim to verify our conclusions using a larger databases constituting many speakers.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP16K01293.

References

- (1) Shohei Kato, etc. "A Preliminary Study of Speech Prosody-Based Relationship with HDS-R Scores Toward Early Detection of Cognitive Impairment in Elderly Using Speech Prosody:Toward Early Detection of Cognitive Impairment in Elderly Using Speech Prosody" JSAI Vol.26, No.2,SP-H pp.347-352 (2011)
- (2) Taler.V, etc. "Comprehensionog grammatical and emotional prosody is impaired in Alzheimer's disease". Neuropsychology. Vol22,No2,pp188-195 (2008)
- (3) Yuto Tanaka, Hiroaki Igaue, Mitsunori Mizumachi and Yoshihisa Nakatoh, "Study of Improvement of Intelligibility for the Elderly Speech Based on Formant Frequency Shift," International Journal of Computer, Consumer and Control (IJ3C), Vol. 3, No.3, pp.57-65, (2014)
- (4) Yumi Wakita, Yuta Yoshida, Mayu Nakamura, "Influence of Personal Characteristics on Nonverbal Information for Estimating Communication Smoothness", proceedings of 18th International conference HCI international 2016 pp.148-157 (2016.7)
- (5) Yumi Wakita, Shunpei Matsumoto, "Communication Smoothness Estimation using F0 Information" Proceedings of the 4th IIAE International Conference on Intelligent Systems and Image Processing 2016 (2016.9)

- (6) Nishio, S., Koyama, K.: A Criterion for Facial Expression of Laugh Based on Temporal Difference of Eye and Mouth Movement. IEICE Vol.J80-ANo.8, 1316-1318 (1997)