# Subjective Age Estimation using Speech Sounds: Comparison with Facial Images

Mayuka Nishimoto Yasuhiro Azuma Naoyuki Miyamoto Takashi X. Fujisawa Noriko Nagata
School of Science and Technology
Kwansei Gakuin University
Sanda, Hyogo 669-1337, Japan
m_nishimoto@ kwansei.ac.jp

***Abstract*— We have defined the perception of one's own age as "subjective age", and have so far approached using other people's facial images. In this paper, we propose a relative estimation method for subjective age by using people's speech sounds and their chronological age. A relative estimation task is performed, wherein an estimator gives rating values to other people about whether they are older or younger than the estimator. In this task, the difference in the actual age between the estimator and the rating value of the person are measured. We plot the results on a two-dimensional plane with the x-axis as the relative age and the y-axis is the estimation result. Thus, the distribution with the plotted points of upper-right direction is obtained, which is approximated by a logistic function. The zero crossing point in the approximation curve with the x-axis is defined as the shift value in the subjective age. 57 total estimators, including 28 males and 29 females from 25 to 44 years old participated in this experiment. As a result, the subjective age using speech sounds tended to be older than using facial images. The tendency was also more remarkably visible in the male groups than in the female group. However, variance was relatively higher in the male young-middle (35-44) and the female young (25-34) groups than using facial images, which indicated that the subjective age varies according to the profile of estimators, such as age and gender.**

***Keywords*— subjective age perception, non-linear regression analysis, face-to-face communication, affective computing**

## I. INTRODUCTION

In face-to-face communication, people estimate the attributes of other people, such as age and gender, simply by looking at them and listening to their voices. In social interactions, age is one of the most important factors, and it assumes even greater significance in the case of first meetings. People subconsciously estimate whether the other person is older or younger, and try to respond in a befitting manner [1].

It has been observed that people generally tend to find other people to appear older than their actual age, while they find themselves to appear younger than their actual age. It can be said that they did not estimate the ages of the other people incorrectly, but that they simply found themselves to appear younger or older than they really were. This perception of one's own age has been termed "subjective age" by the authors. Thus far, we have studied age estimation by using facial images [2–4], and we observed that the subjective age is generally lower than the actual age.

In this paper, we propose a relative estimation method for subjective age by using people's speech sounds and their chronological age. We intend to study the mechanism of age estimation in face-to-face communication by comparing the subjective ages obtained from facial images with those obtained from speech sounds.

## II. PREVIOUS WORKS

Several researches on the perception and estimation of age have been carried out by using facial images. One of the research topics is the accuracy of age estimation. For example, the existence of an "own age/race bias," [5] which implies that people have a superior ability to recognize the faces of people belonging to their own age/race group, has been indicated. In addition, it has been reported that the accuracy of age estimation by elders is lower than that by young people. In addition, it has been reported that the accuracy of age estimation by elderly people is lower than that by young people, which suggests that the accuracy of age estimation is not dependent on age.

It must be noted that in all these researches, the observers exhibited a tendency to estimate the people whose facial images were shown as being older than their actual ages. However, the problem lay not in how the observers perceived the images, but rather in how they perceived their own ages: the observers did not find the people in the images to be older than the observers themselves; instead, they found that they themselves appeared younger than their actual age [2–4]. Thus, although there have been a number of studies on age estimation and perception, the mechanism of age estimation has not yet been clarified.

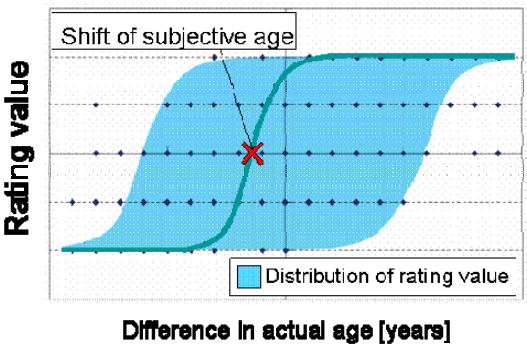
In contrast, previous research has shown that listeners can estimate a talker's age quite accurately based on the talker's speech sounds alone [6]. In particular, it has been indicated that the speech rate plays an important role in age estimation [7-8]. In addition, the perception of a person purely on the basis of his/her voice is strongly related to social factors. It is generally believed that the higher the pitch of voice, the more feminine and docile is the speaker; in addition, a person with a high-pitched voice has been construed as being less selfish and weak [9]. A report has indicated that in Europe as well as in Japan,

people with high-pitched voices are generally perceived as being more child-like and immature. In sum, it can be said that people with high-pitched voices are perceived to be young, while people with low-pitched voices are perceived to be older. Thus, it is evident that the mechanism of age perception has hitherto not been suitably clarified.

The objective of our study is to clarify the mechanism of age perception on the basis of subjective age studies by using not only speech sounds but facial images as well.

## III. CONCEPT OF SUBJECTIVE AGE

In this paper, "subjective age" is defined as one's own age. Figure 1 shows the concept of subjective age and provides an outline of the method for estimating the subjective age. A relative estimation task is performed, wherein an estimator gives rating values to other people who communicate with each other about whether they are older or younger than the estimator. In this task, the difference in the actual age between the person and the estimator and the rating value of the person are measured. As shown in Fig. 1, the data corresponding to the above two parameters are plotted on a two-dimensional plane; the relative age (the chronological age as determined from the speech sound minus the chronological age of the estimator) is plotted on the x-axis, while the estimation result is plotted on the y-axis. If the rating is correct, the data will be distributed in the first quadrant around the origin. If the data is distributed more along the positive direction of the x-axis, it implies that the estimator generally tends to evaluate other people as being older than they really are. On the other hand, if the data is distributed more along the negative x-axis, it implies that the estimator tends to evaluate the other people as being younger than they actually are. It is very difficult to predict the actual age correctly because one generally tends to think of oneself as being older or younger than one actually is. We call this imagined age "subjective age." The shift in the subjective age is the distance between the center of the distribution and the origin.




## IV. EXPERIMENT OF SUBJECTIVE AGE ESTIMATION USING SPEECH SOUNDS

### A. *Speech sounds database*

Speech sounds spoken by 183 female and male speakers from 5 to 76 years old, evenly distributed, have been recorded. Recorded words are the following:

(1) Five vowels in Japanese

(2) Five kinds of greetings:

- "Ohayo" (Good morning)
- "Konnichiwa"(Good afternoon)
- "Tadaima" (I'm home)
- "Arigato" (Thank you)
- "Sayounara" (Goodbye)

Each word, which the speakers spoke with two kinds of facial expressions (with straight face and smiling face), was recorded using a high-quality microphone (SONY ECM-MS95) and a 48 kHz sampling rate.

In our previous work, the rating experiment of subjective age using facial images, we divided estimators into three age ranges (25-34: young, 30.2(O)-7.4(h)0(a)-9(y).9(da9( )6 1-5.8( )-12(y)1-9.3( 30.0.069m(c

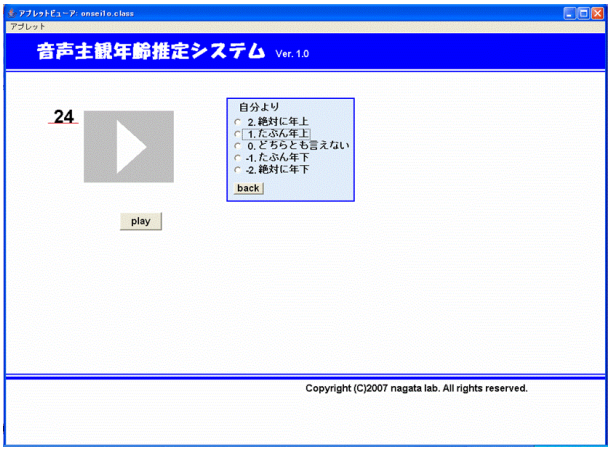



Figure 3. An example of the choosing screen of the subjective age estimation system.

Next, we experimented with a rating scale for these speech sounds, which stimulate the estimators. The estimators were shown 42 speech sounds chosen in the above way with headphones, and they evaluated whether they sounded older or younger than themselves. The evaluation had 5 ranks; "Definitely older than myself (2)," "Probably older than myself (1)," "Not able to estimate (0)," "Probably younger than myself (-1)," and "Definitely younger than myself (-2)." The reason for adopting a range of responses was not to estimate the chronological ages of the speech sounds, but to seek their relative position to others.

Figure 3 shows an example of the chooser screen of the subjective age estimation system, which is constructed by Java. Before starting their experience, estimators were directed to make their decisions speedy, based on intuition, and not to mind the speech sounds of a same person.

### C. *Subjective age quantification*

The subjective ages are calculated to quantify the result of the rating experiments. We plot the results on a two-dimensional plane with the x-axis as the relative age (the chronological age of the facial image minus the chronological age of the subject), and the y-axis is the estimation result. The x-axis ranges from –15 to 15 because the estimators evaluated the speech data with a 15-year difference as a maximum. Thus, the distribution with 42 plotted points of upper-right direction is obtained, which shows the subjective age of the estimator.

Assuming that this distribution is approximated by a logistic function, non-linear regression analysis is applied to the distribution of each estimator. The logistic function here, which converges to the rating values ▪ 2, is defined by the mathematical formula:

$$y = \frac{4}{1+\exp(-a(x-b))} - 2 \qquad (1)$$

Where $a$ is the slope of the curve, $b$ is the zero crossing point in the approximation curve with the x-axis. $a$ and $b$ are estimated by non-linear regression analysis, and $b$ is defined as the shift value in the subjective age. The difference between the actual age of the estimator and the obtained shift value can be defined as the subjective age.

## V. RESULTS

Using the method mentioned above, the rating experiment was conducted and the shift value in the subjective age was calculated on the basis of the data obtained from each estimator.

We eliminated data whose multiple coefficient of determination of the regression curve approximated for each estimator was extremely low ($R^2$<0.10) and, finally, processed the data of 57 total individuals, including 28 males and 29 females.

Two-way ANOVA was executed using gender and age group as independent variables and their shift value in subjective age $b$ calculated by each estimator by regression coefficient as the dependent variable. As a result, a significant gender × age group interaction existed ($F(1, 53) = 4.36$, $p < .05$). The simple main effect test was conducted on the interaction between gender and age groups, where the significant difference was confirmed. Simple main effects of male were significant. The results revealed that the simple main effects of male were significant for age group ($F(1, 53) = 4.11$, $p < .05$). Figure 4 shows the experimental results.

Figure 4. Shift values in the subjective age using speech sounds

Figure 5. Shift values in the subjective age between speech sounds and facial images

In addition, a t-test was conducted on speech sounds and the results of subjective age estimation for each of 4 groups divided by gender and age, as shown in Figure 5. A significant difference between young (25-34) and young-middle (35-44) at male, and young (25-34) at female was confirmed (in order, $t(33) = 2.38$, $p < .05$; $t(43) = 2.64$, $p < .05$; $t(31)=2.70$, $p<.05$, respectively).

First, the subjective age using speech sounds tended to be older than using facial images. To be more precise, in the three groups, except for the female young-middle (35-44) group, the subjective age using speech data obtained a few years higher than using facial images. The tendency was also more remarkably visible in the male groups than in the female group. However, variance was relatively higher in the male young-middle (35-44) and the female young (25-34) groups than using facial images, which indicated that the subjective age varies according to the profile of estimators, such as age and gender.

## VI. Discussions

The subjective age using speech sounds tended to be older than using facial images. Generally, the subjective age in females was older than that in males, which was confirmed to be the same as using facial images. One reason for the higher variance was because this speech data had less information than facial images. Thus, the subjective age using speech sounds was expected to show different interesting properties in social interactions from that of using facial images.

Notably, the estimation accuracy of subjective age improved by eliminating the outliers of speech sounds on the basis of rating data.

## VII. Conclusions

We have proposed an estimation method of subjective age using speech sounds and have compared it with the results from using facial images. Additionally, we have built up a speech-sounds database. The subjective age using speech sounds generally tended to be older than using facial images.

The subjective age represents the relative position to others in a society. Such research becomes more and more important from the viewpoint of self-recognition and socio-psychological effects in the field of HCI, cognitive science, and e-learning.

In the future, we will analyze factors that affect subjective age perception regarding properties of estimators. We are also considering discussing the difference between the voice of one's own and of others and the difference between facial expressions. Ultimately, we will analyze the mechanism of age perception using audio-visual contents, which will be combined with speech and facial data. The experiment will be similar to that of a practical situation of face-to-face communication.

## REFERENCES


[1] Yousuke Arai, and Yutaka Matsui, "The relationship of behavior and affection, to that of the comparison standard which occurs within hierarchical-treatment of those who are of close age," Japanese Journal of Interpersonal and Social Psychology , vol. 31, no. 3, pp. 23-28, 2003.

[2] Noriko Nagata, and Seiji Inokuchi, "Subjective age obtained from facial images –How old we feel compared to others?," V. Palade, R. J. Howlette, and L. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems II, Lecture Notes in Artificial Intelligence 2774, Springer-Verlag, pp. 877-881, 2003.

[3] Naoyuki Miyamoto, Yumi Jinnouchi, Takashi X. Fujisawa, Noriko Nagata, and Seiji Inokuchi, "Estimation of One's Subjective Age Using Facial Images," The IEICE TRANSACTIONS on Fundamentals of Elecronics, Communications and Computer Sciences(Japanese Edition)-A, vol. 90, pp.240–247, 2007.

[4] Takashi X. Fujisawa, Naoyuki Miyamoto, Noriko Nagata, and Seiji Inokuchi, "Estimating one's own subjective age using facial images: An investigation of the factors determining the perception of ourselves as younger," Journal of Japanese Academy of Facial Studies, vol. 7, no.1, pp.121–127, 2007.

[5] Jeffrey S. Anastasi, and Matthew G. Rhodes, "An own-age bias in face recognition for children and older adults," Psychonomic Bulletin & Review, vol. 12, pp.1043–1047, 2005.

[6] D. R. R. Smith, and R. D. .Patterson, "The interaction of glottal-pulse rate and vocaltract length in judgements of speaker size, sex, and age.," The Journal of the Acoustical Society of America, vol. 118, pp. 3177, 2005.

[7] Yumiko Ohara, "Japanese pitch from a sociophonetic perspective," Youko Ide (Ed.), "Women's languages in the world", pp. 42-58, 1997.

[8] Loredana Cerrato, Mauro Falcone, and Andrea Paoloni, "Subjective age estimation of telephonic voices," Speech Communication, vol. 31, no. 2-3, pp. 107-112, 2000.

[9] R. Winkler, "Influences of pitch and speech rate on the perception of age from voice," In: Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, pp. 1849-1852, 2007.