2019 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP2019)
Honolulu, Hawaii, USA, March 4-7, 2019

Study on Perception of Speaker Age by Semantic Differential Method

Yang Li, Maori Kobayashi and Masato Akagi

School of Advanced Science and Technology, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, 923-1292, JAPAN
E-mail: Email: { liyang, maori-k, akagi }@jaist.ac.jp

Abstract

Humans can perceive ages of speakers from uttered voices by their own judgements. The perceived ages are called perceptual ages (PAs). Many earlier studies focused on statistical correlations between aging voices and acoustic features without taking into account the fact that human perception is vague rather than precise [1]. This paper focuses on the psychological factors to study human perceptions for aging voices. An experiment was carried out to evaluate the aging voices by candidates of semantic primitives, and the results of the listening test were analyzed by Semantic Differential Method and Regression Analysis to investigate impressions that human use to estimate PAs of speakers. Results show that with regards to both male and female voices, the Metal Factor (Deep - Flimsy, Full - Delicate, Rich – Thin, Heavy - Light), which shows a linear relation with both male and female PAs, is the most important factor that helps listeners judge PAs of uttered voices. In addition, the rest of the factors show both linear and non-linear relationships with male aging voices, while only non-linear relations with female aging voices.

1. Introduction

In conversations, linguistic information is surely important. However, even without understanding of one language, ages of speakers can still be judged. The judged ages by listeners are called PAs. PAs are useful in our daily life. For example, when answering calls from strangers, listeners always change their reactions and attitudes according to the PAs of the callers.

There have been many reports that studied PAs from engineering aspects. Minematsu et. al (2002) constructed a two-layer-model (acoustic features vs PAs) for estimating human PAs utilizing GMMs. The identification rate was improved to 95% in the subjective elderly group [2]. Yet only two categories, which were subjective elderly (SE) and non-SE (NSE), were too rough for PA estimation. Moreover, the constructed two-layer-model focused on investigating the direct relations between acoustic features and PAs, but paid few attentions to the psychological factors. Against the disadvantages of the two-layer-model, Huang and Akagi (2008) proposed a three-layer model in which people perceive expressive speech not directly from a change of acoustic features, but rather from a composite of different types of “smaller” perceptions that were expressed by semantic primitives (SPs) [1].

Inspired by their study, we assume that human perception for aging voices are leaded by specific SPs. Hence, based on our assumption, we focus on semantic primitives in the second layer and PAs in the top layer in the three-layered model to discuss impressions that human use to estimate PAs of speakers.

2. Listening Tests

2.1 Experiment-I to label PAs

Experiment-I was conducted to label the stimuli with PA. The stimuli were used in formal experiment to evaluate human perception by candidates of semantic primitives.

2.1.1 Databases and stimuli

Listeners can make correct age estimates within ±10 years [3], which means actual ages affect PAs in a sense. Accordingly, for the purpose that the labeled PAs of the stimuli can evenly distribute among the whole age axis, three databases, JNAS [4] (Japanese News Article Sentences), Senior-JNAS [5] and a database of Children's speech [6], were used. The first one gives us speech samples of 153 adult male and 153 adult female speakers of age ranging from 20 to 60, and the second one is composed of speech samples of 151 male speakers and 150 female speakers. Their ages vary from 60 to 90. Speech samples of 6 to 12-year-old boys and girls are included in the third one. The number of samples of boys and girls was 145 and 143, respectively. The stimuli we used were selected from the three databases randomly so as to avoid the effects of subjective factors. It was reported that female PAs are easier to predict [7], from which we can assume that PAs of female are predicted from fewer impressions. In order to verify our assumption, it is needed to study male and female voices separately.

Table 1 shows number of stimuli selected from the three

Table1: Number of stimuli selected from three databases

Actual Age / Gender	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79
Male	15	15	15	15	11	5	15	15
Female	15	15	15	15	15	5	15	15

Table2: Candidates of sematic primitives

Bright-Dark	High-Low	Penetrating-Stuffy
Quiet-Noisy	Violent-Mild	Calm-Restless
Slow-Quick	Warm-Cold	Loud- Tender
Full-Delicate	Soft-Hard	Heavy-Light
Resonant-Non resonant	Lively-Lifeless	Definite-Dull
Clear-Turbid	Tension-Listless	Pleasant-Unpleasant
Beautiful-Ugly	Natural-Unnatural	Smooth-Rough
High extending-High choked	Weakly- Powerful	Articulate-Slurred
Low extending-Low choked	Breathy-Not breathy	Thick-Flimsy
Fine grained-Rough	Rich-Thin	Rounded-Sharp

databases according to their actual age.

2.1.2 Procedure of experiment-I

10 native Japanese in their early 20s (6 males and 4 females) were participated in the experiment. The listening test was done in a quiet room by using a headphone with a sound pressure level of 60dB. One utterance with one sentence per speaker was presented to the subjects and he/she was required to estimate age of speaker in a range of 0 to 100 by a unit of 1 year.

2.2 Experiment-II

Experiment-II was carried out to evaluate human perception for the PAs by candidates of semantic primitives.

2.2.1 Stimuli

For both male and female voices, the stimuli were divided into 8 groups (0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79) according to their PAs labeled in experiment-I. From each group, 5 stimuli, whose value of standard deviation between subjects were the lowest, were used as stimuli.

2.2.2 Candidates of Semantic Primitives

30 pairs of adjectives from the previous studies (Inoue et. al [8], Ueda [9], Nieboer. G et. al [10], Kuriyagawa et. al [11], Tankubo et. al [12], Kido [13]) were selected as candidates. Details of adjectives are shown in Table 2.

2.2.3 Procedure of experiment-II

20 native Japanese who are in their early 20s (16 males and 4 females) were participated in the experiment. The subjects were asked to rate each of the adjectives on a 7-point-scale (as shown in Fig 1) when they heard each utterance, according to intensity of the descriptors they perceived.

Table3: Results of factor analysis (Male Voices)

Pattern Matrix[a]

Semantic Primitives	Factor					
	Mental Factor	Resonance Factor	Power Factor	Beauty Factor	Comfort Factor	Breath Factor
Thick-Flimsy	0.852					
Full-Delicate	0.823					
Rich-Thin	0.720					
Heavy-Light	0.501					
Resonant-Non resonant		0.819				
High-extending-High chocked		0.615				
Listless-Tension		-0.524				
High-Low		0.511				
Clear-Turbid		0.475				
Low extending-Low chocked		0.456				
Lively-Lifeless			0.852			
Bright-Dark			0.703			
Poweful-Weakly			0.622			
Definate-Dull			0.593			
Loud-Tender			0.479			
Violent-Mild			0.438			
Pleasant-Unpleasant				0.687		
Natural-Unnatural				0.628		
Smooth-Rough				0.606		
Articulate-Slurred				0.508		
Beautiful-Ugly				0.497		
Quiet-Noisy				0.460		
Soft-Hard					0.618	
Warm-Cold					0.609	
Sharp-Rounded					-0.533	
Fine grained-Rough						0.616
Breathy-No breathy						-0.498
Cumulative %	22.879	37.333	46.313	53.926	58.383	62.412

Table4: Results of factor analysis (Female Voices)

Pattern Matrix[a]

Semantic Primitives	Factor					
	Mental Factor	Clarity Factor	Active Factor	Power Factor	Comfort Factor	Renonance Factor
Thick-Flimsy	0.854					
Rich-Thin	0.852					
Full-Delicate	0.769					
Loud-Tender	0.459					
Heavy-Light	0.432					
Clear-Turbid		0.740				
Breathy-No breathy		-0.705				
Fine grained-Rough		0.629				
Smooth-Rough		0.547				
Penetrating-Stuffy		0.412				
Definate-Dull		0.410				
Quiet-Noisy			-0.776			
Calm-Restless			-0.601			
Bright-Dark			0.566			
Lively-Lifeless			0.507			
Soft-Hard				-0.731		
Violent-Mild				0.663		
Sharp-Rounded				0.571		
Warm-Cold					0.892	
Pleasant-Unpleasant					0.808	
High-extending-High chocked						0.665
Low extending-Low chocked						0.537
Resonant-Non resonant						0.527
Cumulative %	20.439	34.611	44.588	52.559	57.863	62.946

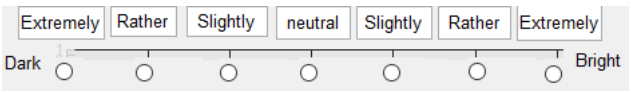

Fig1: Example of 7-point-scale

2.2.4 Analysis in experiment-II

Semantic Differential Method was applied to analyze results of experiment-II. Principle Axis Factoring method was used to extract factors. Eigenvalues, whose value was greater than 1, were regarded as factors. Promax was used as Rotation method. In terms of male voices, 27 pairs of adjectives were classified, while in terms of female voices, 23 pairs of adjectives were classified.

In terms of male voices, results are shown in Table 3.

3. Analysis of Factors

We assume variation of factors are determined by PA so as to investigate whether each factor has relation with PA or not by Regression Analysis. Hence, PAs were regarded as independent variable, while factor score obtained by factor analysis was regarded as the dependent variable.

3.1 Male Voices

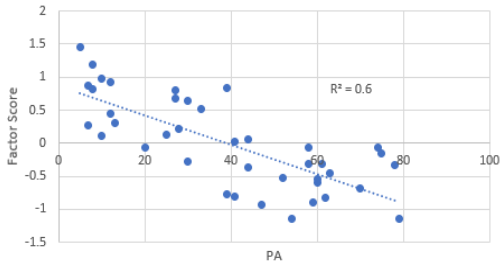

Fig2: Relation between mean value of factor score and PA (Mental Factor)

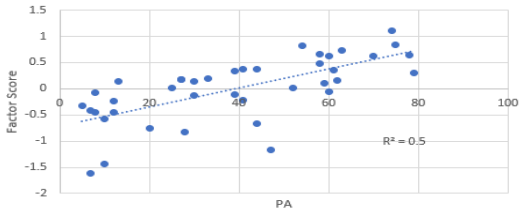

Fig3: Relation between mean value of factor score and PA (Resonance Factor)

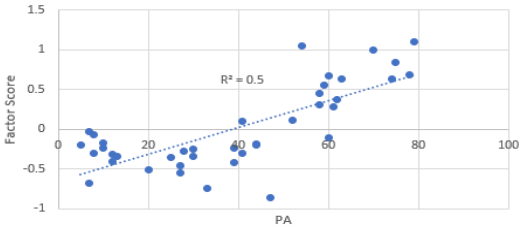

Fig4: Relation between mean value of factor score and PA (Breath Factor)

By regression analyses for male voices, we observe that:

(1) Mental Factor, Resonance Factor and Breath Factor show very strong linear relations with PA. With increasing of PA, male voices become deeper, more listless and breathier, which can be seen from Figures 2, 3 and 4, respectively.

(2) Beauty Factor and Comfort Factor show weak non-linear relations with PA. PA of 40 years old is a peak while under 40 years old, beauty of voices is increase but comfort of voices decrease according to growth of PA. Yet, the trend is reversed while older than 40 years old.

(3) Power Factor has no relation with PA, which may be affected by individual differences the most.

3.2 Female Voices

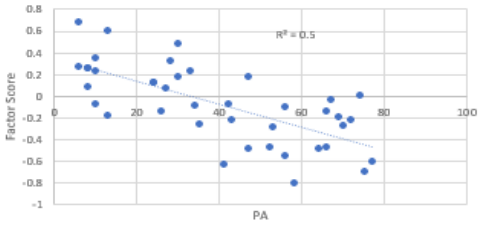

Fig5: Relation between mean value of factor score and PA (Mental Factor)

By regression analyses for female voices, we observe that:

(1) Only Mental Factor show strong linear relations with PA. With increasing of PA, male voices become deeper, which can be seen in Figure 5.

(2) Power Factor and Comfort Factor show weak non-linear relations with PAs. 40 years old is a peak while under 40 years old, power of voices increase but comfort of voices decrease according to growth of PA. Yet, the trend is reversed while older than 40 years old.

(3) Clarity Factor, Active Factor, and Resonance Factor have no relations with PA, which may be affected by individual differences.

4. Verification of Assumptions

For the purpose of verifying whether human perceptions for the aging voices are judged by different semantic primitives, Multiple Regression Analysis is used. Here, we assume that PAs are leaded by combinations of related factors which have strong linear relationship with PAs.

4.1 Male Voices

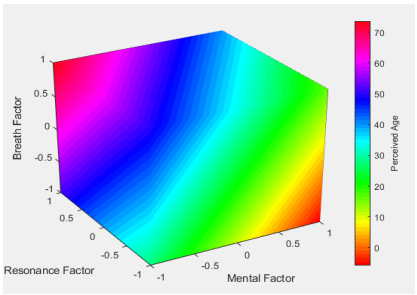

Fig6: Relations between PA of Male and Factor Dimensional Space

Male PAs have strong linear relations with Mental Factor, Resonance Factor and Breath Factor. The constructed factor dimensional space is shown in Fig6.

The equation between PA and the factors is represented as: PA = -16.772 * Mental Factor + 11.889 * Resonance Factor + 11.057 * Breath Factor + 33.91
($R = 0.9$, $R^2 = 0.8$)

4.2 Female Voices

Female PAs have linear relation with Mental Factor. The relation between PA and Mental Factor is shown in Fig7.

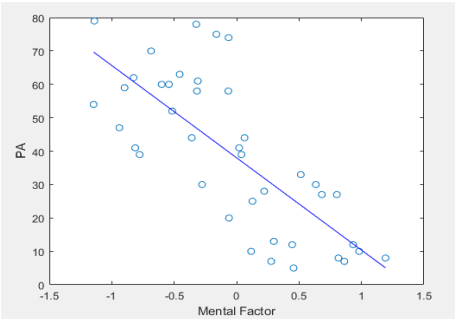

Fig7: Relation between PA and Mental Factor

The equation between PA and Mental Factors is represented as: PA = -27.64 * Mental Factor + 38.01.
($R = 0.6$, $R^2 = 0.5$)

5. Conclusion

This paper focuses on the psychological factors to study human perceptions for aging voices. Two experiments were carried out to evaluate the aging voices by candidates of the semantic primitives, and the results of the experiment were analyzed by Semantic Differential Method and Regression Analysis. As the results:

(1) While Considering semantic primitives determined by PAs, in terms of male voices, PAs have linear relations with Mental Factor, Resonance Factor and Breath factor. Besides, PAs have weak non-linear relations with Beauty Factor and Comfort Factor. Yet for female voices, PAs only have strong linear relation with Mental Factor and weak non-linear relations with Power Factor and Comfort Factor.

(2) On the contrary, while considering PAs are leaded by factors which show linear relations with PAs, in terms of male voices, PAs are leaded by the combination of Mental Factor, Resonance Factor and Breath Factor. While for female voices, PAs are merely leaded by Mental Factor.

(3) Female PAs are easier to predict [7], the reason is explained as female PAs are predicted from fewer impressions.

References

[1] C. F. Huang and M. Akagi, “A three-layered model for expressive speech perception”, Speech Commun 2008., vol.50, pp. 810–828.

[2] N. Minematsu, M. and Sekiguchi and K. Hirose, “Automatic Estimation of One's Age with His/Her Speech Based Upon Acoustic Modeling Techniques of Speakers”, ICASSP 2002, pp.137-140.

[3] M. Huckvale and A. Webb, “A Comparison of Human and Machine Estimation of Speaker Age”, Statistical Language and Speech Processing 2015, pp. 111-112.

[4] The Speech Database Committee, the Acoustical Society of Japan, “ASJ Japanese Newspaper Article Sentences Read Speech Corpus (JNAS)”, http://research.nii.ac.jp/src/en/JNAS.html

[5] K. Shikano, Nara Institute of Science and Technology, “Japanese Newspaper Article Sentences Read Speech Corpus of the Aged (S-JNAS)”, http://research.nii.ac.jp/src/en/S-JNAS.html

[6] Center for Integrated Acoustic Information Research (CIAIR), Nagoya University, “CIAIR Video game Command Voice (CIAIR-VCV)”, http://research.nii.ac.jp/src/en/CIAIR-VCV.html

[7] J. D. Harnsberger, R. Shrivastav, and W. S. Brown Jr., “Modeling perceived vocal age in American English”, Interspeech 2010, pp. 466–469

[8] M. Inoue, T. Kobayasi, “Research field by SD method in Japan and an overview of its adjective versus scale configuration”, Educational Psychology Research 1985, vol. 33, pp. 253−260. (Written in Japanese)

[9] K. Ueda, “Should we assume a hierarchical structure for adjectives describing timbre”, Acoust. Sci. & Tech. 1988, vol. 44, no. 2, pp. 102-107. (Written in Japanese)

[10] G. Nieboer, T. DeGraaf, H. Schutte. “Esophageal voice quality judgements by means of the semantic differential”, J Phonet 1988, pp. 417-36.

[11] M. Kuriyagawa and H. Yahiro, “7 properties of sound quality evaluation”, Acoust. Sci. & Tec1978. Vol. 34, pp. 102-107. (Written in Japanese)

[12] H. Tankubo, K. Kobayashi, “Study on the vocabulary of timbre”, Study of Yuge College 2007. Vol. 29, pp. 57-63. (Written in Japanese)

[13] H. kido, H. Kasuya, “Extraction of common words from daily communication”, Acoust. Sci. & Tech 1999. Vol. 55, pp. 405-411. (Written in Japanese)