

Effects of audiovisual expression of emotion on age perception

Sumi SHIGENO¹

Aoyama Gakuin University, Japan

ABSTRACT

We examined the effects of emotional expression (neutral emotion, happiness, and sadness) on a speaker's age perception in visual and audiovisual condition. Although many researchers have explored the effect of aging on cognitive mechanisms, the effects have not been explored in the view of modalities. In auditory perception on age, Shigeno (1) revealed that a happy (cheerful) voice makes a speaker sound younger. On the other hand, in visual perception of age, the results are not consistent; a smiling face photo makes a speaker look younger (2) or look older (3). The discrepant findings among researchers might be because facial expression photographs capture just the moment of emotional climax of a face. The present study used emotional videos instead of photographs. Participants were asked to identify the ages of twenty-four Japanese native speakers, whose ages were 24–75 years. We compared the identification results between two conditions (visual-only and audiovisual) to examine whether audiovisual happy expressions give younger impression to perceivers than neutral emotion or sadness. The results showed that facial expression does not increase or degrade the effect of vocal emotion. We discussed the integration process of vocal and facial emotion in age perception.

Keywords: Age perception, Audiovisual, Younger or older

1. INTRODUCTION

In our daily communication, a speaker's voice is a cue as important as facial information for estimating his/her age. We often become aware of a speaker's emotions only from his/her voice, as for instance in a telephone conversation. Moreover, we often do not look at a speaker's face well even in a face-to-face situation (4). Therefore, it is surprising that the effect of emotional voices on speakers' age perception remains uninvestigated, except in Shigeno (1), although many researchers have explored the effect of aging on cognitive mechanisms (visual/auditory/multimodal): age differences on cross-modal emotional matching and identification (5); age-related effects on emotion recognition (6); speakers' perceived ages through reading voice (7); subjective age estimation of telephone voices (8); the accuracy of estimates of speaker age (9); the influences of speech rate and speech spontaneity on estimation of speaker age (10).

Shigeno (1) tested the effect of emotional voices on speakers' age perception and found that when a speaker spoke with a happy voice, participants estimated the speaker's age to be younger than his/her chronological age. It was also found that this tendency in female happy voices was stronger than in male happy voices. When a speaker spoke with a sad voice, participants estimated the speaker to be older than his/her real ages. In the perception of vocal emotion, emotional information is mainly defined by pitch shifts of vowels (11, 12). An acoustical speech analysis of the emotional voices indicated that F0 was correlated with the pleasantness–unpleasantness dimension in a two-dimensional psychological space, when calculated by Multi-Dimensional Scaling (MDS) (13). Happiness had higher F0 than other emotions (11, 13). The F0 of younger people is higher than that of older people (particularly for women) (14). Considering these findings, the fact that a happy voice has a higher F0 is the most likely factor to provide a younger impression.

It is generally believed that a smiling face makes a speaker look younger. Voelkle, Ebner, Lindenberger, & Riediger (15) found that facial expressions had a substantial impact on the accuracy and bias of age estimation and relative to other facial expressions, neutral faces were estimated most accurately, while

¹ sshigeno@ephs.aoyama.ac.jp

happy faces were most likely to be underestimated. By contrast, Ganel (3) reported that across different experimental conditions and stimulus sets, smiling faces were consistently perceived as older compared to the same persons' neutral faces. Ganel (3) suggested that Voelkle et al.'s findings are due to observer failure to ignore smile-associated wrinkles, mainly along the region of the eyes. However, Wang et al. (2) and Hass et al. (16) again demonstrated that happy faces look younger and supported the results of Voelke et al. (15). The discrepant findings among researchers might be because a photograph captures just a certain moment of the emotional movement of a face. If the photograph captures the expression at an emotional climax, wrinkles of the face will be deepest. Considering these points, it is necessary to examine the perception of aging using video recordings of facial movement, instead of facial expression photographs, will be necessary.

So far, the effects of speaker's emotion on age perception are discrepant between audition and vision. In our daily lives, we usually estimate the ages of others by both their facial expression and vocal emotion. In speech perception, the visual information of articulatory movements influences the perception of speech: For example, participants presented with a discrepant syllable (an acoustic /ba/ with a visual /ga/) may believe that the speaker has uttered a fused syllable /da/ (17). The present study investigates the process of integration between facial expression and vocal emotion in the audiovisual perception of speaker's age.

In the present study, instead of using photographs of emotional faces, as Ganel (3) and other researchers did, we used videos of facial expression movements. Video gives a more precise impression of aging than photographs do, because with aging, the muscles gradually straighten and shorten, structural aging becomes visible, and with time, the expression of the subject becomes more rigid (18).

2. METHOD

2.1 Participants

The participants with no known hearing disorders participated in the experiment for a partial course credit. In the VO condition, one hundred ninety-three participants ($M_{age} = 20.4$, $SD = 1.2$) were divided into three groups: neutral emotion, happiness, and sadness; each consisted of 59, 68, and 66 participants. In the AV condition, sixty different participants from the VO condition ($M_{age} = 20.7$, $SD = 0.7$) were divided into three groups: neutral emotions, happiness, and sadness; each group consisted of 20 participants.

The participants provided written informed consent. The present study was approved by the Research Ethics Review Board of the College of Education, Psychology and Human Studies at Aoyama Gakuin University, Japan. Participants had previously received hearing tests at university health checkups.

2.2 Stimuli

The stimuli were the sample of emotional faces provided by 24 native Japanese actors, 24–75 years old, recruited from an agency and recorded on a MiniDV tape (SONY, ME DVM60). Each generation (where generation is defined as a decade in the series 20–70) comprised four actors (two of each gender) to avoid age-specific tendencies. The ages of the four actors in the current experiment were around the middle of each generation: the ranges were 24–26, 34–35, 45, 54–56, 64–66, 74–75 years.

The speakers spoke short sentences in Japanese, namely "*Hontou desu ka. Shinji rare masen,*" which can be translated as "Is that true? I can't believe it," in English. These sentences were selected because we often hear them spoken in a happy context (e.g., "He told me about my promotion...") and in a sad context (e.g., "I heard of his death..."). Each speaker spoke them while expressing a neutral emotion, happiness, and sadness, in turn. The speakers were required to use voice pitch ($F0$) in their vocal expressions of emotion, because this is one of the most important parameters of emotional voices (e.g., 9,11), although they could use other expressive features, such as speaking speed and/or loudness of voice. The same sample of speakers was used for both the visual-only experiment and the AV condition. Figure 1 shows three examples of emotional video scenes: neutral, happy, and sad.

The utterances appeared in random order. One session consisted of 72 trials: 3 emotions (neutral/happy/sad) \times 1 verbal stimulus ("Is that true? I can't believe it.") \times 24 speakers.

2.3 Procedure

In the VO condition, the emotional videos without voices appeared in random order. The participants watched the speakers' faces carefully and were judged the speakers' ages. Sessions were conducted for every emotion (neutral/happy/sad); each session consisted of 24 trials: one expression ("Is that true? I can't believe it.") \times 24 speakers. This is because, as Ganel (3) showed, an experimental design that includes repeated presentations of faces of belonging to the same individuals bearing different expressions may bias overt age evaluations. Participants' age evaluations would be biased through their conformity with the commonly held conviction that smiling faces are perceived as younger. To avoid this possibility, the present

design did not use multiple presentations of facial expressions belonging to the same person during the same experimental session.

In the AV condition, on the other hand, the participants watched the speakers' faces and listened to their voices carefully, focusing on the speakers' ages. The other presentation conditions of stimuli were the same as in the VO condition.

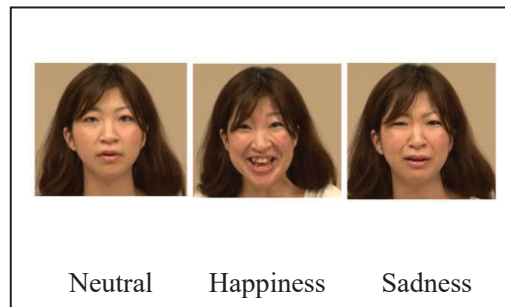


Figure 1 – Three emotional faces of one of the speakers . Scene examples from video recordings

3. RESULTS

3.1 VO condition

To confirm whether the emotions of facial expression movements without voices were perceived as separate emotions (neutral emotion, happiness, or sadness), precedent to the age experiment, emotional facial expressions without voices were identified by 24 other participants (4 males, $M = 21.5$, $SD = 1.29$; 20 females, $M = 21.5$, $SD = 4.15$). The averaged percentages of correct identification were as follows: neutral emotion, $M = 56.4$, $SD = 21.2$; happiness $M = 94.6$, $SD = 12.6$; sadness, $M = 47.4$, $SD = 14.8$. It was recognized that neutral emotion and sadness were difficult to identify only based on the emotional faces (Table 1).

The results of age judgments were averaged across all age groups and across both genders. Figure 2 shows the average of the speakers' perceived age as a function of emotion and speakers' age. A two-way mixed-type ANOVA (emotion [neutral, happiness, and sadness] and speaker's age [20s, 30s, 40s, 50s, 60s, and 70s]) was conducted on the estimated ages. The effect of emotion was not significant, $F(2, 190) = 0.033$, $p = .968$, $\eta^2 = .000$. A significant effect of age was found, $F(2.562, 486.799) = 1633.96$, $p < .001$, $\eta^2 = .896$. There was no significant interaction, $F(5.124, 486.799) = 1.859$, $p = .098$, $\eta^2 = .019$. The results confirmed that younger speakers were perceived as younger and older speakers were perceived as older.

The results did not support the results of Ganel (3), who indicated that the photos of smiling faces look older than those of neutral faces, nor the opposite results that the photos of smiling faces look younger than those of neutral faces (2, 15, 16). Instead, the present results indicated that there were no differences in perceived ages among three facial emotions: neutral emotion, happiness, and sadness.

Table 1 – Average percentages of identification of facial emotion and standard deviations

Facial Emotion		Perceived Emotion					
		Neutral	Happiness	Anger	Disgust	Fear	Sadness
Neutral	M	56.4	3.8	17.9	16.0	2.6	3.3
	SD	21.2	7.3	13.8	12.8	3.2	3.5
Happiness	M	2.4	94.6	0.7	1.0	0.5	0.7
	SD	5.3	12.6	2.4	4.3	1.9	1.6
Sadness	M	5.2	3.6	13.5	16.7	13.5	47.4
	SD	6.6	6.3	10.7	10.5	10.1	14.8

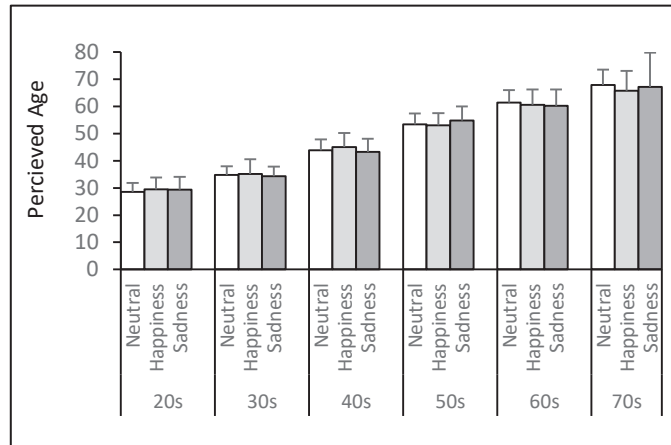


Figure 2 – Perceived age of speakers when the emotions were presented visually

3.2 AV condition

To confirm whether the emotions of audiovisual emotional stimuli were perceived as a particular emotion (neutral emotion, happiness, or sadness), precedent to the age experiment, emotional audiovisual expressions were identified by the other 21 participants (4 males, Mage = 24.3, SD = 3.86; 17 females, Mage = 21.7, SD = 4.47). Averaged percentages of correct identification were the follows: neutral emotion, M = 53.4, SD = 18.5; happiness M = 95.6, SD = 4.8; sadness, M = 63.5, SD = 13.3. It was recognized that neutral emotion and sadness were not easy to identify based on the audiovisual emotional stimuli (Table 2).

Table 2 – Average percentages of identification of audiovisual emotion and standard deviations

Audiovisual		Neutral	Happiness	Anger	Disgust	Fear	Sadness
Emotion							
Neutral	M	53.4	0.6	21.4	20.6	2.0	2.0
	SD	18.5	1.5	16.5	15.5	2.5	3.1
Happiness	M	3.2	95.6	0.2	0.2	0.4	0.4
	SD	3.9	4.8	0.9	0.9	1.3	1.3
Sadness	M	1.6	2.4	5.6	8.5	18.5	63.5
	SD	2.1	2.1	5.8	8.3	10.8	13.3

Figure 3 shows the average of the speakers' perceived ages as a function of emotion and speakers' ages. A two-way mixed-type ANOVA (emotion [neutral, happiness, and sadness] and speaker's age [20s, 30s, 40s, 50s, 60s, and 70s]) was conducted on the estimated ages. The effect of emotion was not significant, $F(2, 57) = 0.282, p = .755, \eta^2 = .010$. A significant effect of age was found, $F(2.512, 143.173) = 1328.991, p < .001, \eta^2 = .959$. There was no significant interaction, $F(5.024, 143.173) = 1.955, p = .089, \eta^2 = .064$.

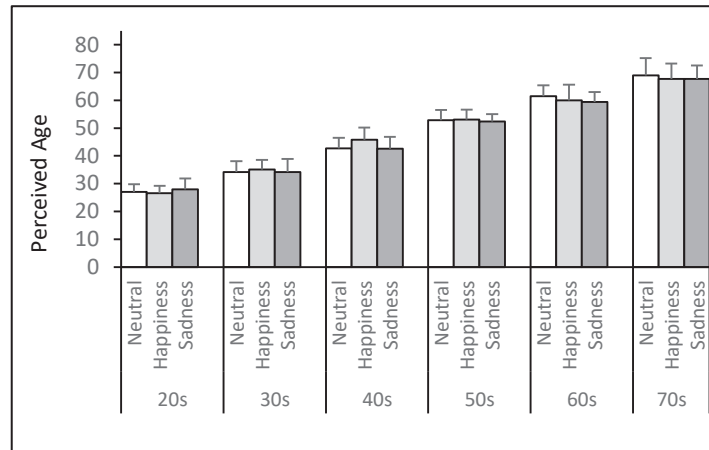


Figure 3 – Perceived age of speakers when the emotions were presented audiovisually

4. DISCUSSION

Shigeno (1) indicated that happy voices sound younger than their chronological ages when only voices are heard. In contrast, the current results did not indicate that either smiling faces or smiling faces with happy voices will help the perceivers have impressions that the speakers are older or younger than their ages. The results suggest that facial expression might dampen or offset the effect of vocal emotion in age perception.

In speech perception, we integrate the gesture of speech production into what we hear (17), which occurs robustly even when subjects are aware of an incompatibility between the auditory and visual components of speech. The current study suggests that the integration process of a speaker's emotions between vision and audition in the perception of speaker's age might be different from that of speech perception.

5. CONCLUSIONS

This study indicates that although a happy (cheerful) voice makes our impression younger, happy facial movements do not make us look younger. Furthermore, when we look at someone speaking with a happy voice and with a happy face, our age recognition are made often on the basis of the speaker's facial expressive movements. As noted by Moyse (19), although voices are often considered to be the auditory counterparts of faces, the comparison between voice and face is not always obvious; methods and dependent variables of age estimation research differed between the studies of facial expression and of vocal expression. Further research is necessary to elucidate the discrepancies in age perception between facial expression and vocal emotion and the mechanism of the audiovisual integration process between facial expression and vocal emotion.

ACKNOWLEDGEMENTS

This research was supported by the Japan Society for the Promotion of Science under Grant Number 24530919.

REFERENCES

1. Shigeno, S. (2016). Speaking with a happy voice makes you sound younger. *International Journal of Psychological Studies*, 8(4), 71-76. doi:10.5539/ijps.v8n4p71
2. Wang, Z., He, X., & Liu, F. (2015). Examining the effect of smile intensity on age perceptions. *Psychological Reports: Mental & Physical Health*
3. Ganel, T. (2015). Smiling makes you look older. *Psychonomic Bulletin & Review*, 22(6), 1671–1677. doi: 10.3758/s13423-015-0822-7
4. Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial clues*. Englewood Cliffs, NJ: Prentice-Hall.

5. Hunter, E. M., Phillips, L. H., & MacPherson, S. E. (2010). Effects of Age on Cross-Modal Emotion Perception. *Psychology and Aging, 25*(4), 779–787. <http://dx.doi.org/10.1037/a002052>
6. Chaby, L., Luherne-du Boullay, V., Chetouani, M., & Plaza, M. (2015). Compensating for age limits through emotional crossmodal integration. *Frontiers in Psychology, 6*, 1–12. doi: 10.3389/fpsyg.2015.00691
7. Ptacek, P. H., & Sander, E. K. (1966). Age recognition from voice. *Journal of Speech, Language, and Hearing Research, 9*(2), 273–277. <http://dx.doi.org/10.1044/jshr.0902.273>
8. Cerrato, L., Falcone, M., & Paoloni, A. (2000). Subjective age estimation of telephone voices. *Speech Communication, 31*(2), 107–112. [http://dx.doi.org/10.1016/S0167-6393\(99\)00071-0](http://dx.doi.org/10.1016/S0167-6393(99)00071-0)
9. Eriksson, E., Green, J., Sjöström, M., Sullivan, K. P., & Zetterholm, E. (2004). Perceived age: a distracter for voice disguise and speaker identification? *Proceedings of FONETIK 2004*, 80–83.
10. Waller, S. S., Eriksson, M., & Sörqvist, P. (2015). Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. *Frontiers in psychology, 6*, 1–11. doi: 10.3389/fpsyg.2015.00978
11. Murray, I. R. and Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America, 93*(2), 1097–1108. doi: 10.1121/1.405558
12. Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614–636. <http://dx.doi.org/10.1037/0022-3514.70.3.614>
13. Shigeno, S. (2004). Recognition of vocal expression of emotion and its acoustic attributes. *Japanese Psychological Research, 74*(6), 540–546. doi.org/10.4992/jjpsy.74.540
14. Russell, A., Penny, L., & Pemberton, C. (1995). Speaking Fundamental Frequency Changes Over Time in Women: A Longitudinal Study. *Journal of Speech, Language, and Hearing Research, 38*(1), 101–109. doi:10.1044/jshr.3801.101
15. Voelkle, M. C., Ebner, N. C., Lindenberger, U., & Riediger, M. (2012). Let me guess how old you are: Effects of age, gender, and facial expression on perceptions of age. *Psychology and Aging, 27*(2), 265–277. <http://dx.doi.org/10.1037/a0025065>
16. Hass, N., Weston, T. D., & Lim, Seung-Lark. (2016). Be happy not sad for your youth: the effect of emotional expression on age perception. *PloS one*, 1-13. doi:10.1371/journal.pone.0152093
17. McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746 – 748. doi:10.1038/264746a0
18. Le Louarn, C. (2009). Muscular aging and its involvement in facial aging: the Face Recurve concept. *Annales de dermatologie et de venerologie. 136*(4), S67-72. doi: 10.1016/S0151-9638(09)74530-d
19. Moyse, E. (2014). Age estimation from faces and voices: A review. *Psychologica Belgica, 54*(3), 255–265. <http://dx.doi.org/10.5334/pb.aq>