Speaking with a Happy Voice Makes You Sound Younger

Sumi Shigeno[1]

[1] Department of Psychology, Aoyama Gakuin University, Tokyo, Japan

Correspondence: Sumi Shigeno, Department of Psychology, Aoyama Gakuin University, Tokyo, Shibuya, 150-8366, Japan. Tel: 81-3-3409-7653. E-mail: sshigeno@ephs.aoyama.ac.jp

Abstract

This study investigates the effects of emotional voices (expressing neutral emotion, sadness, and happiness) on a judgement of a speaker's age. An experiment was conducted to explore whether happy voices sound younger than neutral and sad voices. The identification of 24 speakers' ages (12 of each gender) based on their emotional voices was done by 40 participants. The speakers' ages were 24-75 years. Participants identified the age of each speaker only by hearing his/her emotional voice. The results showed that when a speaker spoke with a happy voice, participants estimated their age to be younger than their chronological age. Furthermore, the results regarding female happy voices were more conspicuous than male happy voices. In contrast, when a speaker spoke with a sad voice, participants estimated them to be older. The results suggest that a happy voice sounds younger because of its higher voice pitch (*F0*). We discussed the role of vocal pitch and other paralinguistic factors for providing an aging impression.

Keywords: happy voice, speaker's age, voice pitch, younger

1. Introduction

It is generally believed that a smiling face makes a speaker look younger. In fact, most models for cosmetic advertisements smile to highlight the cosmetics' effect of giving a younger appearance. However, as indicated by Ganel (2015), only one study has examined the effects of smiling on perceived age (Voelkle, Ebner, Lindenberger, & Riediger, 2012). Voelkle et al. (2012) found that age estimation ability decreased with age; furthermore, facial expressions had a substantial impact on the accuracy and bias of age estimation and relative to other facial expressions, the age of neutral faces was estimated most accurately while the age of faces displaying happy expressions was most likely underestimated. However, in contrast to their findings, Ganel (2015) reported that across different experimental conditions and stimulus sets, smiling faces were consistently perceived as older compared to the same persons' neutral faces. Ganel (2015) suggested that the effect reported by Voelkle et al. (2012) is due to observer failure to ignore smile-associated wrinkles, mainly along the region of the eyes.

In our daily communication, a speaker's voice is an important cue to estimate his/her age as much as facial information. We often understand a speaker's emotions only from his/her voices, as exemplified through a telephonic conversation. Moreover, we often fail to look at a speaker's face even in a face-to-face situation (Ekman & Friesen, 1975). By doing so, the information regarding a speaker's emotion would be received through auditory and visual senses. Therefore, it is surprising that the effect of emotional voices on speakers' age perception remains unexplored, although most researchers have explored the effect of aging on cognitive mechanisms: age differences on cross-modal emotional matching and identification (Hunter, Phillips, & MacPherson, 2010); age-related effects on emotion recognition (Chaby Luherne-du Boullay, Chetouani, & Plaza, 2015); speakers' perceived ages with reading voice (Ptacek & Sander, 1966); subjective age estimation of telephone voices (Cerrato, Falcone, & Paoloni, 2000); the accuracy of estimates of speaker age (Eriksson, Green, Sjöstrom, Sullivan, & Zetterholm, 2004); influences of speech rate and speech spontaneity on estimation of speaker age (Waller, Eriksson, & Sörqvist, 2015).

Can emotional voices influence our recognition of a speaker's age? More specifically, can we assume that someone who is speaking in a happy voice sounds younger than his/her actual age? Spoken language has various paralinguistic information (e.g., emotional state, age, and gender), besides lexical meanings. Emotion is one of the most essential sources which helps us know the situation of the conversation; for example, people speaking

with a happy voice to convey intimacy. This study aimed to test the influence of happy voices on aging impression.

Considering that facial expression influences the vocal emotion perception (Shigeno, 1998), the experimental settings used in this study were only auditory so that participants could focus on emotional voices. In an age test, participants were required to estimate the age of speakers, whose chronological ages were between 24 and 75 years, only by hearing their emotional voices (expressing neutral emotions, sadness, or happiness).

2. Method

2.1 Participants

Forty participants (5 males, M_{age} = 22.0, SD = 0.71; 35 females, M_{age} = 21.4, SD = 0.81) were asked to estimate the age of the speakers by hearing their emotional voices. None had any known history of hearing deficits. 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.

2.2 Stimuli

The sample of emotional voices was provided by 24 native Japanese actors, aged 24-75, recruited from an agency and recorded on a ProRes recorder (AJA, KI-PRO). Each generation (20-70) comprised four actors (two of each gender) to avoid age-specific tendencies. For example, Torre III and Barlow (2009) observed greater variability in speech productions by older adults compared with their younger, same-sex counterparts, as evidenced by larger standard deviations for their measures. 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, such as "Hontoo desu ka. Shinji rare masen", which corresponds to the sentences "Is that real? I can't believe it" in English. These sentences were selected because we often hear these 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 neutral emotion, happiness, or sadness. The speakers were required to use voice pitch ($F0$) in their vocal expressions of emotion because it is one of the most important parameters of emotional voices (e.g., Murray & Arnott, 1993; Shigeno, 2004), although they were allowed to use expressions such as speaking speed and/or loudness of voice. As a result, when they expressed the emotions in a sad tone, their voices were a little lower and softer than a happy emotion. The speakers were also instructed not to use other elements, such as laughing, crying, or clicking their tongues because these elements had not been sufficiently studied to define individual differences among speakers. The speakers practiced their expression of emotions several times before their voices were recorded more than twice. The noiseless recordings that two researchers specializing in cognitive psychology (the experimenter and a doctoral student) judged to be the most emotional were selected as stimuli.

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

2.3 Procedure

Forty participants carefully heard the speakers' voices and focused on the speakers' ages. They estimated and wrote down each speaker's age. All conditions of the stimuli were met by the participants. A 200-ms pure tone of 1000 Hz was presented 2.0 s before each utterance as a warning. Participants were given 4 s to record each judgment.

The voice recordings were played from the PC (HP, ProBook 650G1) through loudspeakers (BOSE, 301V; 101 MM) at a level comfortable for listening, so that the listening environment would not differ much from daily life. The experiment was conducted in a quiet room.

3. Results

First, the emotional voices were identified by other 58 participants (15 males, M_{age} = 21.3, SD = 0.90; 43 females, M_{age} = 21.0, SD = 0.65) to confirm whether the emotions of voices were perceived as a separate emotion (neutral, happiness, or sadness). Table 1 shows the results. A two-way repeated measures analysis of variance (ANOVA) according to the factors of emotion (neutral vs. happiness vs. sadness) and gender (male vs. female speakers) was conducted on correct (i.e., intended) emotion identification. There was a significant main effect of emotion, $F(2, 114)$ = 137.49, p < .001, η^2 = .71. Multiple comparisons with Bonferroni correction revealed that the most correctly identified emotion was happiness (p < .001) and the least correctly identified emotion was neutral (p < .001). There was no significant main effect of gender, $F(1, 57)$ = 0.001, p = .975, η^2 = .00. There was a significant interaction between emotion and gender, $F(2, 114)$ = 55.39, p < .001, η^2 = .49. A further analysis was

conducted to explore the effect of vocal emotion in each gender. As a result, the effect of vocal emotion was significant: male speakers, $F(2, 78) = 40.21$, $p < .001$, $\eta^2 = .51$ and female speakers, $F(2, 78) = 61.92$, $p < .001$, $\eta^2 = .61$. Multiple comparisons with Bonferroni correction between any two emotions were significant ($p < .001$) except between neutral and sadness expressed by male speakers.

Table 1. Average percentages of identification of vocal emotion

Vocal emotion		Neutral	Response Happiness	Anger	Disgust	Fear	Sadness
Neutral	Male	54.2	0.9	8.9	23.3	2.3	10.5
	Female	34.6	0.6	29.7	26.9	2.0	6.2
Happiness	Male	13.3	80.9	0.8	3.0	1.3	0.8
	Female	1.4	86.4	1.9	3.2	3.7	3.4
Sadness	Male	4.0	5.6	1.0	2.3	33.8	53.3
	Female	0.6	5.8	7.5	5.8	13.5	67.0

Since our emotional lives have richness and diversity (Barrett, 2009), the identification percentages were scattered across the six emotion categories although the participants were required to identify voices expressing three emotions (neutral, happiness, and sadness). Emotion perception studies have reported that a speaker's intended expression is often identified as different emotions (Shigeno, 1998); identification rates differ greatly depending on the types of emotion (Banse & Scherer, 1996); and neutral emotional voices are identified as specific emotions (Liu & Pell, 2012). In fact, in the present results, neutral voices were perceived less properly than happiness and sadness. For example, neutral female voices were identified as anger (29.7%) and disgust (26.9%) although the percentage of identification as neutral (34.6%) was identified most among the six alternatives. On the other hand, it was more clearly indicated that happiness and sadness were identified as respective emotion most.
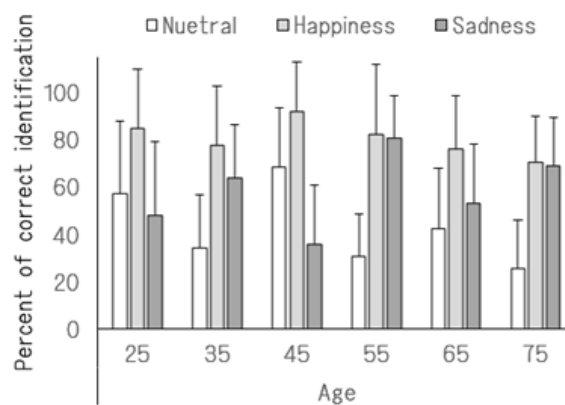

Figure 1. Averaged correct percentages of perceived emotions as a function of speakers' age

Figure 1 shows the correct (i.e., intended) identification of each emotion by each age group. A two-way repeated measures ANOVA according to the factors of emotions (neutral vs. happiness vs. sadness) and an averaged speaker's age (25, 35, 45, 55, 65, and 75) was conducted for correct (i.e., intended) emotion identification. Significant main effects of emotion and the speaker's age were obtained: $F(2, 114) = 114.96$, $p < .001$, $\eta^2 = .67$; $F(5, 285) = 7.32$, $p < .001$, $\eta^2 = .114$. There was a significant interaction between emotion and averaged speaker's age, $F(10, 570) = 15.25$, $p < .001$, $\eta^2 = .44$.

The results of age judgments were then averaged across all age groups and for both genders. The happy voice of one male speaker in his 50s was excluded from the calculation as it was not perfectly presented. Significant differences were obtained between any two emotions and between male and female speakers.

Figure 2 shows the average of the speakers' estimated age as a function of emotion. A two-way repeated measures ANOVA according to the factors of emotion (neutral vs. happiness vs. sadness) and gender (male vs. female speakers) was conducted on the estimated ages. A significant main effect of emotion was found, $F(2, 78)$ = 58.45, $p < .001$, $\eta^2 = .60$. Multiple comparisons with Bonferroni correction revealed that the perceived age of happy voice was younger than perceived age of neutral ($p < .001$) and sad voices ($p < .001$). No significant main effect of gender was found, $F(1, 39) = 3.42$, $p = .072$, $\eta^2 = .08$. There was a significant interaction, $F(2, 78)$ = 30.31, $p < .001$, $\eta^2 = .44$, showing that the influence of the three vocal emotions on age identification differs between male and female speakers. As the interaction was significant, further analysis was conducted to explore the effect of vocal emotion in each gender. The simple main effect of vocal emotion was significant for male speakers, $F(2, 78) = 40.21$, $p < .001$, $\eta^2 = .51$; for female speakers, $F(2, 78) = 61.92$, $p < .001$, $\eta^2 = .61$, showing that perceived age of voices differed among the three emotions as indicated in Figure 2. Multiple comparisons with Bonferroni correction between any two emotions expressed by either male or female voices were significant ($p < .001$) except between neutral and sadness expressed by male speakers. Thus, happy voices were judged to be younger than neutral and sad voices and sad voices were judged to be older than neutral (except in female speakers) and happy voices. Furthermore, the current results indicate that, in relation to happiness and sadness, the perceived female age was younger than the perceived male age, happiness: $F(1, 39) = 4.33$, $p < .05$, $\eta^2 = .10$; sadness: $F(1, 39) = 23.33$, $p < .001$, $\eta^2 = .37$.
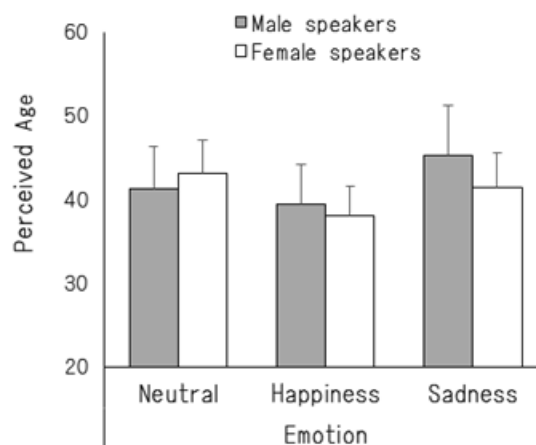

Figure 2. Averaged perceived ages of speakers as a function of neutral, happy, and sad voices. Significant differences were obtained between any two emotions and between male and female speakers

Table 2. Age differences by subtracting chronological age from perceived voice's age

Vocal emotion	Averaged chronological age					
	25	35	45	55	65	75
Neutral	1.1	-4.3	-8.1	-7.2	-12.0	-16.8
Happiness	0.2	-7.1	-12.7	-14.2	-18.4	-15.0
Sadness	-0.3	-2.7	-4.4	-2.8	-16.1	-13.5

Note. Positive age differences indicate that a speaker's voice sounds older; negative age differences indicate that a speaker's voice sounds younger.

Table 2 shows whether the speakers sound younger or older. Age differences were calculated by subtracting chronological age from emotional voice's age. Positive figures indicate that a speaker's voice sounds older than his/her chronological age. In contrast, negative figures indicate that a speaker's voice sounds younger than his/her chronological age. Table 2 indicates that older speakers' voices sound younger than their chronological ages, and the difference was conspicuously larger than the case of younger speakers' voices. A two-way repeated measures ANOVA according to the factors of emotion (neutral vs. happiness vs. sadness) and age (25, 35, 45, 55, 65, and 75) on age differences confirmed the results. A significant main effect of emotion was found, $F(2, 78)$ =

57.56, $p < .001$, η2 = .60. The main effect of age was also found to be significant, $F(5, 195) = 153.27$, $p < .001$, η2 = .80, and multiple comparisons with Bonferroni correction indicated that the identification of the younger speakers (in their twenties) was accurate (i.e., small age differences) ($p < .001$) and the older speakers (in their 60s and in their 70s) were identified as younger than their chronological ages ($p < .001$). It is possible to say that the current participants were in their early 20s (i.e., young) and therefore, the perceived ages of younger speakers were more accurate than those of older speakers. The results suggest that an own-age bias also exists (Moyse, 2014) in the age estimation of emotional voices.

4. Discussion

Although, in the facial age estimation, Ganel (2015) suggested that smiling faces look older, the present study indicated that a happy voice sounds younger in the contexts where only voices are heard. Why does someone speaking with a happy voice sound younger than his/her real age? Emotional information is mainly defined by pitch shifts of vowels (Murray & Arnott, 1993; Banse & Scherer, 1996); an acoustical speech analysis of the emotional voices indicated that $F0$ was correlated with the pleasantness–unpleasantness dimension in a two-dimensional psychological space, as calculated by Multi-Dimensional Scaling (MDS) (Shigeno, 2004); happiness had higher $F0$ than other emotions (Murray & Arnott, 1993; Shigeno, 2004); and the $F0$ of younger people is higher than that of older people (particularly for women) (Russell, Penny, & Pemberton, 1995). Considering these findings, a higher $F0$ of happiness can be the most possible factor that can provide a younger impression.

Using 374 normal and healthy Japanese speakers (187 males and 187 females) from adolescent or older age groups, Nishio and Niimi (2005) reported that changes in Speaking Fundamental Frequency (SFF) associated with aging. They observed that females in their 30s and 40s showed clearly lower frequencies than those in their 20s. Across all age groups, including those in their 80s, SFF tended to decrease with aging. Gelfer and Schofield (2000) reported that subjects perceived as female had a higher mean SFF and higher upper limit of SFF than subjects perceived as male. They further reported that a significant correlation was achieved between upper limit of SFF and ratings of femininity. These investigations could explain the current results that the perceived female age is younger than the perceived male age when they speak in happy or sad voices. However, it remains unexplained that in the case of neutral voices, perceived female age is older than the perceived male age.

On the other hand, with aging, our voice becomes hoarser and more breathy (Gorham-Rowan & Laures-Gore, 2006). Therefore, older people might find it hard to produce a loud and long voice. In fact, when the speakers produced vocal emotions, their voices were different not only in pitch but also in the length and/or loudness of voice. These factors might influence the speaker's age estimation.

In conclusion, a substantial body of present studies suggests that a happy voice sounds younger because of its higher pitch, although during facial recognition, it looked older (Ganel, 2015). The results further showed that the tendency is more conspicuous in female happy voices than male happy voices. As noted by Moyse (2014), 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 studies using faces and those using voices as stimuli. Further research would be necessary to elucidate the discrepant results between male and female speakers and between facial expression and vocal emotion. The results suggest that if a person wishes to give the impression of being younger, he/she should speak with a happy voice.

Acknowledgments

References

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

Barrett, L. F. (2009). Variety in the spice of life: A psychological construction approach to understanding variability in emotion. *Cognition Emotion*, *23*(7), 1284-1306. http://dx.doi.org/10.1080/02699930902985894

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

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. http://dx.doi.org/10.3389/fpsyg.2015.00691

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial clues*. Englewood Cliffs, NJ: Prentice-Hall.

Eriksson, E., Green, J., Sjöstrom, M., Sullivan, K. P., & Zetterholm, E. (2004). Perceived age: A distracter for voice disguise and speaker identification? In *Proceedings of FONETIK 2004* (pp. 80-83).

Ganel, T. (2015). Smiling makes you look older. *Psychonomic Bulletin & Review*, *22*(6), 1671-1677. http://dx.doi.org/10.3758/s13423-015-0822-7

Gelfer, M. P., & Schofield, K. J. (2000). Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as male. *Journal of Voice*, *14*(1), 22-33. http://dx.doi.org/10.1016/S0892-1997(00)80092-2

Gorham-Rowan, M. M., & Laures-Gore, J. (2006). Acoustic-perceptual correlates of voice quality in elderly men and women. *Journal of Communication Disorders*, *39*(3), 171-184. http://dx.doi.org/10.1016/j.jcomdis.2005.11.005

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/a0020528

Liu, P., & Pell, M. D. (2012). Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behavior Research Methods*, *44*(4), 1042-1051. http://dx.doi.org/10.3758/s13428-012-0203-3

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

Murray, I. R., & 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. http://dx.doi.org/10.1121/1.405558

Nishio, M., & Niimi, S. (2005). Changes in speaking fundamental frequency characteristics with aging. *The Japan Journal of Logopedics and Phoniatrics*, *46*(2), 136-144. http://dx.doi.org/10.5112/jjlp.46.136

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

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. http://dx.doi.org/10.1044/jshr.3801.101

Shigeno, S. (1998). Cultural similarities and differences in the recognition of audio-visual speech stimuli. *Proceedings of International Congress on Spoken Language Processing*, *2*, 281-284.

Shigeno, S. (2004). Recognition of vocal expression of emotion and its acoustic attributes. *Japanese Psychological Research*, *74*(6), 540-546. http://dx.doi.org/10.4992/jjpsy.74.540

Torre III, P., & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders*, *42*(5), 324-333. http://dx.doi.org/10.1016/j.jcomdis.2009.03.001

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

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.

Copyrights