

ACOUSTICAL LETTER

Speech levels: Do we talk at the same level as we wish others to and assume they do?

Akira Toyomura[1,2,*], Tetsunoshin Fujii[3], Kazuyo Nakabayashi[4], David R. R. Smith[4], Jun Toyama[5] and Yasuhiro Kawabata[3]

[1]*Graduate School of Health Sciences, Gunma University, 3–39–22 Showa-machi, Maebashi, 371–8514 Japan*
[2]*Department of Human System Science, Hokkaido University, Kita 10, Nishi 7, Kita-ku, Sapporo, 060–0810 Japan*
[3]*Department of Psychology, Hokkaido University, Kita 10, Nishi 7, Kita-ku, Sapporo, 060–0810 Japan*
[4]*Department of Psychology, Faculty of Health Sciences, University of Hull, Hull, HU6 7RX, UK*
[5]*The Institute for Practical Application of Mathematics, Yamanote 1–4–2–18, Nishi-ku, Sapporo, 063–0001 Japan*

(*Received 8 May 2020, Accepted for publication 30 July 2020*)

Keywords: Speech level, Assumption, Preference
PACS number: 43.70.Mn [doi:10.1250/ast.41.841]

1. Introduction

In humans, the ability to control the level of the voice in speech is an important social skill. We speak in a wide range of environments and social situations, with listeners at different distances from us. Most people are aware of the appropriate range of speech levels for each social situation — we have our indoor and outdoor voices. It is important for speakers to control their own vocal level so that their speech is intelligible and appropriate to the listener and social context. A loud voice does not necessarily increase speech intelligibility [1] and, if the speech level is too high, then the listener can form a negative impression of the speaker [2–4]. There is then a pressing need for speakers to regulate and maintain their speech level at the appropriate level across many different types of social and environmental situations.

To achieve this fine balance, speakers should be able to unconsciously estimate how the intended listener would hear their voice. In this sense, it is thought that the speakers have unconscious self-model of their own voice (i.e., an assumption of how their own voice is perceived), which allows them to control their own voice level to suit the environment or social situation they find themselves in. This psychological experience of one's own voice (which we call 'voice sound') can be thought of as being comparable to the 'body image' we have of our own body [5]. Previous studies on body image have not addressed the issue of voice sound. However, we suggest that one's own voice constitutes a part of the 'body image' because the voice is produced by the combined action of the diaphragm, lungs, vocal cords, and various articulators such as the tongue, palate, lips, velum and nasal cavity, and the voice sound is perceived and adjusted through auditory feedback. Although previous studies have investigated how a speaker's speech level affects judgements of his/her personality [2–4], the accuracy of the estimation of the speaker's own voice level remains unexplored. Investigating the accuracy of estimating one's own voice level (voice sound) would shed light on the nature of the self-model (or body image) as it relates to one's own voice.

*e-mail: ak.toyomura@gmail.com

The aim of the present study is twofold. One is to examine how a speaker's own speech level correlates with the level at which the speaker assumes that their listener is experiencing. Another is to understand the speaker's preference for the voice level of received speech (i.e., when others are talking). We hypothesized that the vocal level at which the speaker spoke would be close to the preferred vocal level of received speech.

2. Methods

2.1. Participants

Seventeen Japanese speakers (5 females), with a mean age of 29.1 years ($SD = 8.67$), from Hokkaido University participated in this study. None reported any history of neurological injury or disease, or vocal pathology. All participants passed a hearing screening at 20-dB HL bilaterally at 500, 1,000, 2,000 and 4,000 Hz. Informed consent was orally obtained from all individuals prior to their participation in accordance with the Declaration of Helsinki.

2.2. Procedure

The experiment was conducted in a soundproof room. A loudspeaker (PIONEER S-170II) was used to produce the sound of various speakers' voices. Each participant sat on a chair facing the front of the loudspeaker. The distances between the participant and loudspeaker were 0.5, 1, 2, and 4 m, presented in a random order in each condition. The experimenter changed the distance by moving the loudspeaker. Nothing was placed between the participant and loudspeaker. A volume control for the output of the loudspeaker was positioned next to the participants. The initial volume was set at random levels in each experiment session.

For the hearing task, voices saying "ohayo" ("good morning" in Japanese) from five speakers unknown to the participants were recorded prior to the experiment, with one voice to be used for practice while the other four voices (two from each gender) being used in the experimental trials. The mean fundamental frequencies (F0) of the four voices were 121 Hz and 122 Hz in male voices, and 244 Hz and 208 Hz in female voices. When presenting these voice stimuli, the order of the four voices were randomized between participants. Each participant was given detailed instructions on the

experimental procedure. Practice trails were run to familiarize the participant with the upcoming task.

In the following vocalization and hearing tasks, in order to directly compare the three measured values in the following analysis, the A-weighted, maximum sound pressure levels were symmetrically measured for the loudspeaker and participant: the level of the participants' voice was measured at the side of loudspeaker, and the loudspeaker's output was measured at the side of the participants' ear.

The experiment proceeded in the following steps: first, in order to measure the preferred sound level for each voice, the voices from the four speakers were presented to participants through a loudspeaker from each distance (0.5, 1, 2 and 4 m). Participants were instructed to adjust the volume of each voice coming from the loudspeaker to the level which would be a comfortable listening level if someone were at the loudspeaker position talking to them. The voice was presented repeatedly until participants were satisfied with the level of the speech coming from the loudspeaker. On completion, the maximum sound pressure of the sound was measured at the side of the participant's ear. The mean of the sound pressures of the four voices was calculated for each distance in each participant. This measure is hereafter called the *Preference*.

Next, the participants' normal speech vocal level was recorded. They spoke 'ohayo' three times as they would normally do while assuming that there was a listener at the four different distances of the loudspeaker. The sound pressure of each vocalization was measured at the distance the participant was told to assume their listener was there. After every vocalization the participants were asked to judge whether their voice level was the same level as usual or not. When it was different, they were asked to vocalize again. The mean of sound pressures of three voices was calculated for each distance. This measure is hereafter called the *Vocalization*.

Third, the participants' assumptions about their own vocalization were measured. First, they were instructed to imagine a listener at the loudspeaker position, and to say "Ohayo" repeatedly toward the loudspeaker (the imagined listener). Then, they were instructed to imagine what level the listener at the loudspeaker position would hear. Next, they were instructed to imagine the speaker and listener had changed places; in other words, the listener at the loudspeaker was the self (the participant) and the participant was the listener (the listener sitting in the participant's position). The participant's own voice recorded at each distance condition during the previous procedure (*Vocalization*) was played repeatedly from the loudspeaker. The participant was then instructed to adjust the volume of the loudspeaker to the level which they had just imagined (i.e. the level the listener at the loudspeaker position would hear). If they forgot the imagined volume during the adjustment, they were asked to vocalize again. These instructions were given before this session, but the same explanations were provided during measurement if necessary, until the participants fully understood the procedures. After the adjustment, the sound pressure level of the voice was measured on the side of the participant's ear. This measure is hereafter called the *Assumption*.

3. Results

Figure 1 shows sound pressure levels for the three levels (*Preference*, *Vocalization*, and *Assumption*). A speaker with a quiet voice tended to have low sound pressure levels for the three measure at any distances (e.g., participant 1 in Fig. 1), and a speaker with a loud voice tended to have high sound pressure levels at any distances (e.g., participant 9 in Fig. 1). However, regardless of the vocal levels, the *Vocalizations* tended to have the largest sound pressure among the three levels across the four distances in each participant. A two-way repeated measures ANOVA with distance (0.5 m, 1 m, 2 m, and 4 m) and condition (*Preference*, *Vocalization*, and *Assumption*) as the within-participant factors revealed a main effect of condition ($F(2, 32) = 27.34$, $p < 0.001$, partial $\eta^2 = 0.63$, *Preference*: 52.53 dB, *Vocalization*: 58.39 dB, and *Assumption*: 51.02 dB), suggesting that the voice levels of *Preference*, *Vocalization*, and *Assumption* were significantly different from each other, irrespective of the distances. However, it did not show significant difference between the distances ($F(3, 48) = 0.91$, $p = 0.44$, partial $\eta^2 = 0.05$, 0.5 m: 54.17 dB, 1 m: 54.55 dB, 2 m: 53.54 dB, and 4 m: 53.68 dB), which indicates that the voice levels reaching the measurement point in each condition were not different between distances. There was no significant interaction between condition and distance ($F(6, 96) = 1.61$, $p = 0.15$, partial $\eta^2 = 0.09$).

Pairwise comparisons using *t*-test with Bonferroni correction were conducted in each distance condition (0.5, 1, 2, and 4 m). All distance conditions produced the same patterns of significance: the difference between *Vocalization* and *Preference* was significant (0.5 m: $t(16) = 4.20$, $p < 0.01$, $d = 0.99$, 5.62 dB; 1 m: $t(16) = 6.05$, $p < 0.001$, $d = 1.42$, 7.48 dB; 2 m: $t(16) = 3.12$, $p < 0.05$, $d = 0.88$, 4.41 dB; 4 m: $t(16) = 4.67$, $p < 0.001$, $d = 1.28$, 5.93 dB). The difference between *Vocalization* and *Assumption* was also significant (0.5 m: $t(16) = 6.35$, $p < 0.001$, $d = 1.02$, 6.67 dB; 1 m: $t(16) = 7.54$, $p < 0.001$, $d = 1.30$, 8.67 dB; 2 m: $t(16) = 5.08$, $p < 0.001$, $d = 1.10$, 6.47 dB; 4 m: $t(16) = 6.82$, $p < 0.001$, $d = 1.45$, 7.66 dB). However, the difference between *Preference* and *Assumption* was not significant (0.5 m: $t(16) = 0.80$, $p = 1.00$, $d = 0.15$, 1.04 dB; 1 m: $t(16) = 1.01$, $p = 0.98$, $d = 0.19$, 1.19 dB; 2 m: $t(16) = 1.82$, $p = 0.26$, $d = 0.32$, 2.07 dB; 4 m: $t(16) = 1.23$, $p = 0.71$, $d = 0.30$, 1.73 dB).

To investigate the relation between *Preference*, *Vocalization* and *Assumption*, the means of the sound pressure levels at the four distances (0.5 m, 1 m, 2 m, and 4 m) were tested using a Pearson's correlation coefficient test with Bonferroni correction. Figure 2 shows the relations between the three levels. Significant correlations were observed for *Assumption* vs. *Preference* ($r = 0.78$, $p < 0.001$) and *Vocalization* vs. *Assumption* ($r = 0.82$, $p < 0.001$). There was a non-significant correlation between *Vocalization* and *Preference* under the Bonferroni correction ($r = 0.54$, $p = 0.08$).

The absolute difference values in sound pressure level were calculated for each participant in each distance, and were averaged across the four distances (Fig. 3). The pattern of our results suggest that speakers spoke (V: *Vocalization*) about 7 dB louder than the level at which they prefer to be

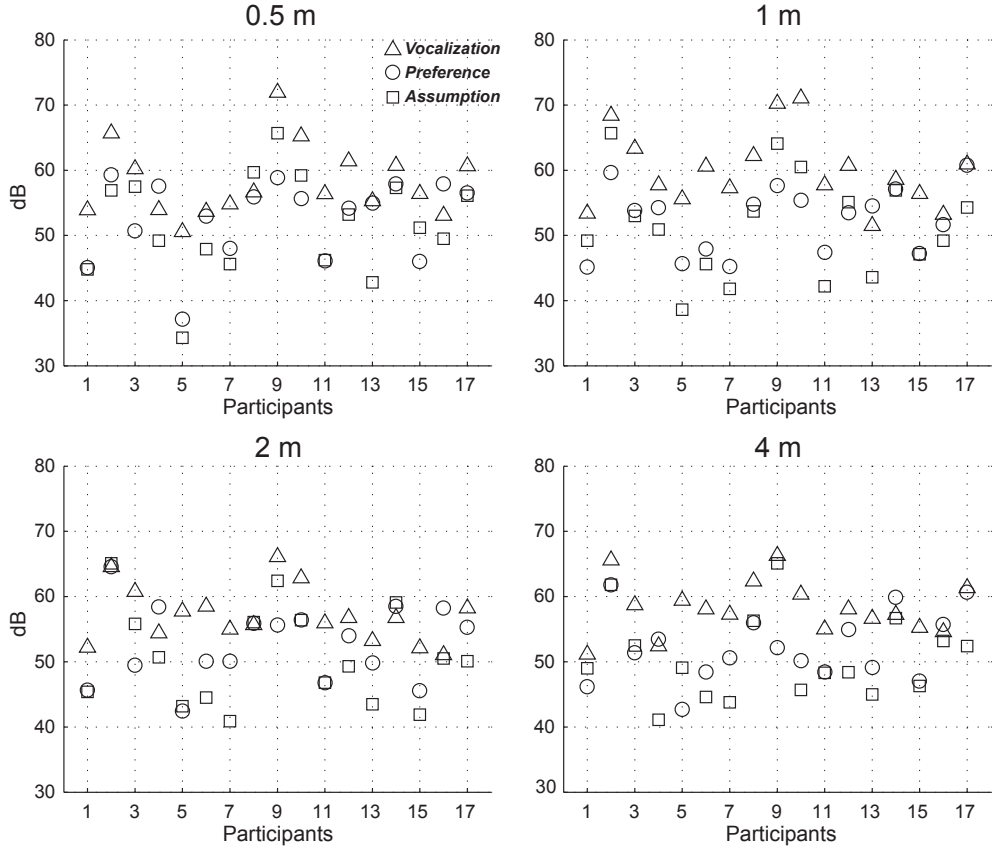

Fig. 1 Sound pressure levels of *Vocalization* (triangle), *Preference* (circle), and *Assumption* (square). Participant number is represented on the *x*-axis, and the *y*-axis represents the sound pressure level. For most participants and distances, the *Vocalizations* tended to show the largest sound pressure among the three levels.

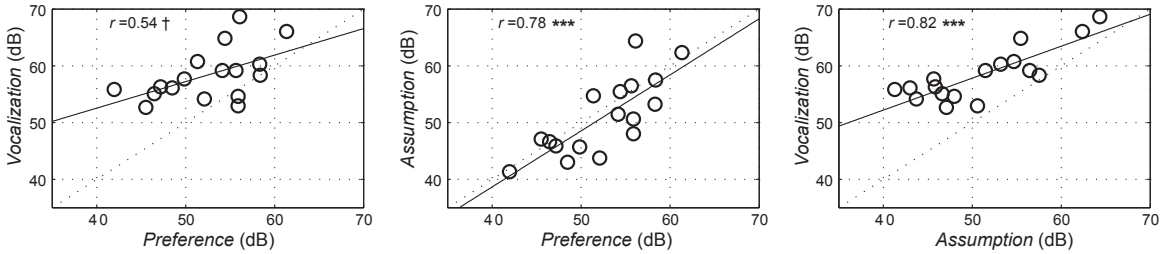

Fig. 2 Relations between *Vocalization*, *Preference*, and *Assumption*. The circles represent means of the four distance conditions for each participant. Relation between *Vocalization* and *Preference* suggests that a speaker who wants to hear a loud voice from others tended to vocalize loudly, and vice versa. $^{***}p < 0.001$, † trend toward significant ($p = 0.08$).

spoken (P: *Preference*) (|V-P|, 6.73 ± 3.51 dB (mean $\pm$ SD)), and at which they assume their voice is heard (A: *Assumption*) (|V-A|, 7.56 ± 3.58 dB). There is a smaller difference of 4.10 (± 2.31) dB in |A-P|. One-way repeated measures ANOVA with factors of absolute difference between vocal levels (|V-P|, |A-P| and |V-A|) showed a significant main effect of difference ($F(2, 32) = 6.89$, $p < 0.01$, partial $\eta^2 = 0.30$). Pairwise comparisons using a *t*-test with Bonferroni correction showed that the |V-A| was significantly greater than the |A-P| ($t(16) = 3.53$, $p < 0.01$, $d = 1.15$). The comparison of the |V-P| and the |A-P| showed a trend toward significance ($t(16) = 2.53$, $p = 0.07$, $d = 0.89$). The |V-P| was not significantly different from the |V-A| ($t(16) = 0.93$, $p = 1.0$, $d = 0.23$).

4. Discussion

Comparing the three sound pressure levels in each participant, the *Vocalization* (speaker's own speech level) was larger than *Preference* (speaker's preference for the voice level of received speech) and *Assumption* (speech level at which the speaker assumes that their listener is experiencing) regardless of the distance. We speculate that there are at least three possible reasons for this result. First, it is known that the middle-ear reflex occurs by stiffening the stapedius and tensor

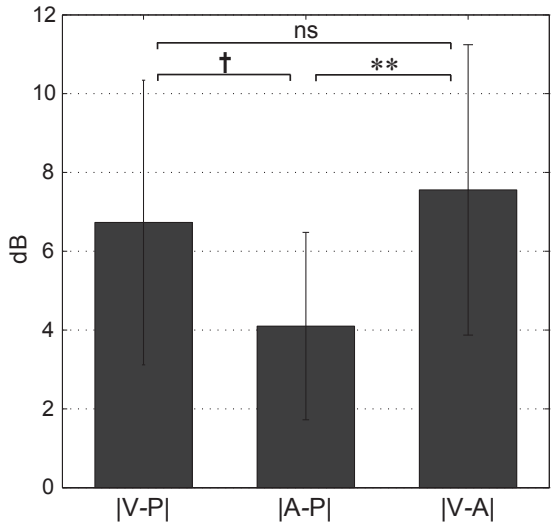

Fig. 3 Mean across all distances and for all participants for the three absolute differences between vocal conditions. Graphs represent means ± SD. $^{**}p < 0.01$, † trend toward significant ($p = 0.07$), ns: not significant.

tympani muscles attached to the ossicles just before speaking, which reduces the effect of self-generated sounds such as speech. Previous studies reported that the reduction effect of the middle-ear reflex is around 1–20 dB [6]. Therefore, that the speakers of this experiment estimated the sound level of their own voice as being less than it was in reality seems entirely understandable. For this reason, it appears that the speakers had a misassumption that they were vocalizing at a level closer to their preferred level, but their real vocalization level was about 7 dB higher than they assumed (|V-A|). Secondly, the high vocal level might be an adaptive strategy learnt through experiences of various social environments. In everyday life, our communications sometimes take place in noisy environments, so we might have learnt to talk louder under such conditions in order to compensate for the reduction in intelligibility. Sound pressure level reduces inversely with the square of the distance between speaker and listener. Hence, to compensate for the distance between ourselves and the listener, we may speak louder than we wish as our voice will be attenuated before it reaches the ear of the listener. A high vocal level might therefore be an adaptive strategy to enable oneself to be heard when communicating with others under a number of social conditions and at different distances. Thirdly, the speech used in this experiment was a single word of greeting and not a string of words conveying a complex sentence. Considering a previous study showing that speakers reduce their vocal level when they produce a complex message [7], the 7 dB difference (|V-A| and |V-P|) found in this experiment might be reduced when participants make a longer speech including messages. In addition, when we say greeting words such as "good morning" used in this experiment, it is necessary to convey (mostly) positive emotions included in the intonation of the voice rather than the content of the speech. Therefore, the large sound pressure level might be used so that the emotion can be more easily delivered.

Another main finding of this study was that the speakers had an assumption that they were vocalizing at a level closer to their preferred level (the *Assumption* was close to the *Preference*). This result may indicate that the speaker unconsciously intends to make the communication smooth, by using a voice with comfortable sound pressure levels. Alternatively, the "preferred" vocal level may be close to "easily understandable" level for the speaker. This assumption of the speaker also could contribute to smooth communication.

5. Conclusions

This study has shown that speakers assume that they are vocalizing to a listener at or close to the speakers' own preference level, but in reality speakers are vocalizing at a level higher than what the speakers assume they are. If one's own voice could be assumed to be part of the 'body image' [5], our findings would suggest that self-model of own voice (the 'voice sound') speakers have is not necessarily accurate in terms of volume. However, this voice sound is perceptually accurate in that it accounts for the natural physics of sound propagation upon the voice image. This suggests that the voice sound, as part of the body image and as an extended part or object ('embodiment') of that body image, is tied to the perception of the receiver as well as the transmitter. The discrepancy found in this study may reduce if participants make a longer speech including messages. This speculation requires additional study.

Acknowledgments

This study was supported by JSPS KAKENHI (Grant Numbers 21700232, 24680025, 24650138, 16K00366 and 19H04195). No conflicts of interest to be declared.

References

[1] S. S. Barreto and K. Z. Ortiz, "Influence of speech rate and loudness on speech intelligibility," *Pró Fono*, **20**, 87–92 (2008).
[2] R. A. Page and J. L. Balloun, "Effect of voice volume on perception of personality," *J. Soc. Psychol.*, **105**, 65–72 (1978).
[3] H. Hollien, M. P. Gelfer and T. Carlson, "Listening preferences for voice types as a function of age," *J. Commun. Disord.*, **24**, 157–171 (1991).
[4] A. C. Nichols and E. M. Muller, "The development of procedures for the measurement of vocal loudness behaviors," *J. Psycholinguist. Res.*, **21**, 41–57 (1992).
[5] T. F. Cash, "Body image: Past, present, and future," *Body Image*, **1**, 1–5 (2004).
[6] R. A. Butler, "The influence of the external and middle ear on auditory discriminations," in *Auditory System*, W. D. Keidel and W. D. Neff, Eds. (Springer, Berlin, 1975).
[7] J. W. Black, "The relation between message type and vocal rate and intensity," *Speech Monogr.*, **16**, 217–220 (1949).