The Effect of Speaker-Specific Auditory Images on Reading in Japanese

Yuichi Mori and Keiko Ishii

Department of Psychology, Graduate School of Humanities, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501

Correspondence should be addressed to:

Keiko Ishii

Graduate School of Humanities, Kobe University

1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501

JAPAN

E-mail: ishii@lit.kobe-u.ac.jp

Phone & Fax: +81-78-803-5518

Yuichi Mori declares that he has no conflict of interest. Keiko Ishii declares that she has no conflict of interest.

Abstract

Auditory images for speech are preserved and can be accessed during reading. This research, conducted in Japan, examined whether and to what extent previous findings on the influence of speaker-specific auditory images in reading can be generalized to non-English speakers in a different cultural context. In two studies, Japanese participants were asked to read a text aloud after being informed that the text had been written by either a fast speaker or a slow speaker whose speech they were to listen to. The participants read the episode more slowly when it was attributed to a slow speaker than when it was attributed to a fast speaker. Individual differences in one's mimicry of the speaker moderated the influence of speaker-specific auditory images in reading. The influence was confirmed only for those who consciously mimicked the speaker. In contrast, situational cues manipulated to generate affiliation with and closeness to the speakers did not influence the participants' reading times.

Keywords: auditory imagery, voice, reading, mimicry, culture

The Effect of Speaker-Specific Auditory Images on Reading in Japanese

In daily communications, listeners usually pay attention to not only the verbal meaning of a spoken message but also the nonlinguistic features of a speaker's voice, such as pitch and loudness. Listeners likely preserve nonlinguistic features in memory because in many cases, the features are useful for detecting speakers' identity, social status, and emotional states (e.g., Nygaard, Burt, & Queen, 2000; Palmeri, Goldinger, & Pisoni, 1993). The nonlinguistic features of a speaker's voice also influence a variety of linguistic tasks (for a review, see Hubbard, 2010). For instance, when participants are asked to listen to messages spoken by a speaker and then read a text written by the speaker, the speaker's speaking rate influences their reading times (Alexander & Nygaard, 2008; Kosslyn & Matt, 1977). In the present research, testing non-English speakers, we examine whether the nonlinguistic features of a speaker's voice influence reading times and whether influence by nonlinguistic features is moderated by individual differences in mimicry of the features.

Researchers have mainly focused on auditory images of basic vocal features and investigated whether and to what extent auditory images are perceptually evoked and accordingly influence one's perceptual experience (e.g., Intons-Peterson, 1980; Intons-Peterson, Russell, & Dressel, 1992). Researchers have also extended the findings by focusing on more complex stimuli such as music (Halpern, 1988; Hubbard & Stoeckig, 1988) and environmental sounds (Stuart & Jones, 1996). Specifically, research on musical auditory imagery has examined brain activity detected in auditory imagery and auditory perception and found that similar areas of temporal lobes were activated in both cases (Zatorre, Halpern, Perry, Meyer, & Evans, 1996), suggesting that auditory imagery involves the same brain areas as auditory perception.

Moreover, whether and to what extent auditory images for speech are preserved and

can be accessed during reading has been examined. Abramson and Goldinger (1997) demonstrated that when participants were asked to make lexical decisions, they took more time to respond to stimuli with long-vowels than to stimuli with short-vowels, even if the stimuli were presented visually and were equated for orthographic length. This finding suggests that participants could access phonological representations which preserved the temporal aspects (e.g., vowel duration) of overt speech, and these representations were activated in their silent reading.

Further evidence suggests that auditory imagery containing perceptual details of speaker-specific information can be accessed during reading. In a study by Kosslyn and Matt (1977), participants were asked to read a text aloud or silently after being informed that the text had been written by either a fast speaker or a slow speaker whose speech they were to listen to. In the reading aloud condition, reading times for the text were longer when the writer was supposed to be a slow speaker than when the writer was a fast speaker, whereas the speaker's speaking rate did not influence the participants' reading times in the silent reading condition. Adopting the same experimental paradigm used by Kosslyn and Matt (1977), Alexander and Nygaard (2008) replicated the findings in the reading aloud condition and reported that the influence of auditory images on reading was more evident in people with higher auditory imagery ability than in those with lower auditory imagery ability. Moreover, Kurby, Magliano, and Rapp (2009) demonstrated that when participants listened to conversations between two speakers and were then asked to read scripts with the speakers, the activation of auditory imagery during reading was enhanced when the participants read familiar scripts with the speakers (Study 1) and had extended experience with the speakers' voices (Study 2). This data suggests that participants may perceptually simulate the speakers' voices while being exposed to

the conversation, and the perceptual simulation (Barsalou, 1999) may accordingly influence their reading.

The present research aims to investigate whether the influence of speaker-specific auditory images in reading (Alexander & Nygaard, 2008; Kosslyn & Matt, 1977) can be generalized to non-English speakers in a different cultural context. For this purpose, we focused on Japanese speakers living in Japan. Reflecting cultural differences in communication practices in terms of the emphasis on contextual information (e.g., Hall, 1976), East Asians, including Japanese speakers, are more likely than Westerners, including North Americans, to pay more spontaneous attention to paralinguistic cues such as vocal tone in the comprehension of speech (e.g., Ishii, Reyes, & Kitayama, 2003; Kitayama & Ishii, 2002; Tanaka et al., 2010). East Asians' higher sensitivity to paralinguistic cues may suggest that they can retain speaker-specific auditory images of nonlinguistic features more easily and are more likely to be influenced by the auditory images during their reading. Moreover, reflecting the culturally dominant value of interdependence (Markus & Kitayama, 1991), compared to Westerners, East Asians are more tuned into the other's perspective in a communication game (Wu & Keysar, 2007). Cultural practices emphasizing interdependence and focus on the other may encourage Japanese readers to simulate perceptually a speaker's voice while they are presented with the speaker's conversation and to read a text as if mimicking how the speaker spoke. Taken together, we expect the influence of speaker-specific auditory images in reading to appear even among Japanese speakers. As the influence of auditory images has not been tested cross-culturally, to our knowledge, we tested only Japanese speakers as an initial step in this research.

This study also aims to explore the conditions under which the influence of speaker-specific auditory images in reading would or would not appear. We considered two

issues to address this point. First, people's likelihood of using auditory images of nonlinguistic features may depend on the extent to which they mimic the speaker. Individual differences may exist in whether people consciously mimic the speaker, and accordingly, the differences may moderate the influence of speaker-specific auditory images in reading. For instance, Alexander and Nygaard (2008) addressed the effect of individual differences in imagery ability and found that when an easy text was presented for reading, the influence of speaker-specific auditory images was greater for participants who are high in imagery ability than for those who are low in imagery ability. While auditory imagery ability corresponds to the vividness of representations of images, the activation of representations may depend on a subjective factor like whether or not participants tried consciously to mimic the speakers. To our knowledge, the effect of the subjective factor has not been fully examined. In the present research, by asking participants whether they tried to mimic the speaker after the reading task, we thus addressed the effect of one's mimicry of speakers directly, instead of imagery ability. We expect the influence of speaker-specific auditory images in reading to be more prominent among participants consciously mimicking the speaker than the other participants.

Second, some situational cues might foster the influence of speaker-specific auditory images in reading, as Kurby et al. (2009) suggested. In this research, we focused on situational cues associated with mimicry. Previous studies have found that mimicry is associated with a desire for affiliation (Lakin & Chartrand, 2003; van Baaren, Horgan, Chartrand, & Dijkmans, 2004) and that a desire for interpersonal orientation fosters spontaneous attention to nonlinguistic cues such as vocal tone (Gray, Ishii, & Ambady, 2011; Ishii & Uchida, 2016). One powerful cue to trigger a desire for affiliation and interpersonal orientation is human faces. There is evidence that exposure to schematic human faces fosters one's interpersonal orientation

(Ishii, Kobayashi, & Kitayama, 2010). In addition, as previous studies have suggested (Stel & Vonk, 2010), individuals' perceived closeness to the speaker might also be associated with their mimicry of the speaker. While Kurby et al. (2009) manipulated perceptual familiarity in terms of script and speakers' voices, the current study focused on more interpersonal features like perceived closeness to speakers. In summary, by presenting participants with schematic human faces in Study 1 and by manipulating the speech content in terms of perceived closeness to speakers in Study 2, we also examined the possibility that these situational cues moderate the influence of speaker-specific auditory images in reading. We expected that participants presented with schematic faces during the experiment (Study 1) and participants feeling a closeness to speakers by listening to a speech with more familiar content (Study 2) would be more influenced by speaker-specific auditory images than the other participants.

Study 1

Method

Participants. Sixty-five Japanese undergraduates (32 men and 33 women, M_{age} = 18.92 years, *SD* = 1.15) participated in this study. All participants were native Japanese speakers. They were tested individually in Japanese and paid 500 yen (about 4 USD).

Materials. All the stimuli were prepared in Japanese. Following the procedure used in Alexander and Nygaard (2008), two conversations were developed by asking four amateur actors (two females and two males) to read a script, which was created based on an everyday conversation between undergraduates. The roles of the fast speaker and slow speaker were counterbalanced between actors of the same gender. Thus, four conversations (2 [speakers' gender] × 2 [speaking rate]) were prepared. Each of the conversations lasted approximately 90 seconds. The mean speaking rate was 9.57 Japanese syllabary characters per one second among

the faster speakers and 7.36 Japanese syllabary characters per one second among the slower speakers. Two texts describing an everyday event among typical undergraduates were also developed. The length of the two episodes was almost identical (206 and 204 characters). Which episode was attributed to either a fast speaker or a slow speaker was also counterbalanced.

Procedure. Participants were randomly assigned to either the control or face condition. Each participant was escorted to an experimental booth and seated in front of a desk where a laptop computer and a microphone were set up. On the computer screen, schematic faces, which were adopted from Ishii et al. (2010), and a laboratory's logo were presented in the face condition as desktop wallpaper, whereas pictures of houses and the same logo were presented in the control condition (see Figure 1). No one was suspicious about this manipulation.

To begin with, participants listened to a brief conversation to familiarize them with two speakers and the speaking rates in the conversation. Approximately half of the participants (16 men and 16 women) listened to a conversation between two male speakers, while the other half (16 men and 17 women) were presented with a conversation between two female speakers. Before this, they were told that they would answer questions related to the conversation. We then asked them eight questions as a check to confirm that they had paid attention to the conversation. The mean accuracy of their answers was 90.4% (*SD* = 12.3), suggesting that, overall, the participants understood the conversation very well.

Participants were then asked to read two episodes aloud in front of the computer and the microphone. They were told that one episode had been written by one of the speakers and the other episode had been written by the other speaker. Prior to reading, for each episode, participants listened for approximately 5 seconds to the first part of the conversation again in which the speaker's name (either Mr. [Ms.] Suzuki or Mr. [Ms.] Sato) was included, so that they

recognized which of the two speakers wrote the episode. It should be noted that participants could correspond the speakers to the writers at that time without receiving any information on the speakers' speaking rates. Participants started reading after a signal from the experimenter and read the entire episode. Reading times were measured in seconds from the signal until they finished reading. Approximately half of the participants were asked to read an episode written by a faster speaker first and then read the other episode by a slower speaker. The order was reversed for the remaining participants.

Finally, participants were asked to fill in a post-questionnaire for the manipulation check. The questionnaire asked, "Did you try to mimic how the speakers had spoken?" The response options were "yes" or "no." Participants were also presented with nine statements related to a speaker's personality ("Mr. [or Ms.] Suzuki [or Sato] is friendly", "is likable", "is cold-hearted", "is similar to you", "is a person who can have a good connection with you", "is a person who can get on well with you", "is a person who can have a deep connection with you", "is a person who has less to do with you", and "is a person with whom you want to interact") and asked to rate to what extent they agree with each statement for each speaker using a 7-point scale (1 = not at all, 7 = completely). We computed the mean rating score for each participant by averaging scores for the nine statements (Cronbach's $\alpha = .88$, "is cold-hearted" and "is a person who has less to do with you" were reversed items) so that speakers were evaluated more positively as the mean rating scores increased.

Results and Discussion

We computed the mean reading times for each participant across episodes written by a fast and a slow speaker. We also checked whether participants consciously mimicked the speakers. Of the 65 participants, 19 of them (9 in the control condition and 10 in the face

condition) reported that they consciously mimicked the speakers, while the remaining 46 participants (24 in the control condition and 22 in the face condition) did not engage in mimicry. The ratio did not differ across the affiliation conditions, $\chi^2(1) = 0.01$, $p = .94$. We then performed an analysis of variance (ANOVA) with three between-subject variables (speakers' gender, affiliation condition, and mimicry) and one within-subject variable (speaking rate). All data analyzed in this study are included as a supplementary information file. Relevant mean reading times are shown in Table 1.

The main effect of the speaking rate was significant, $F(1, 57) = 8.43$, $p = .005$, $\eta^2_p = .13$. Participants took a longer time to read an episode attributed to a slow speaker ($M = 29.84$, $SD = 3.03$) than an episode attributed to a fast speaker ($M = 29.48$, $SD = 2.90$). The interaction between mimicry and speaking rate was also significant, $F(1, 57) = 5.90$, $p = .02$, $\eta^2_p = .09$. Participants reporting that they had consciously mimicked the speakers took a longer time to read an episode attributed to a slow speaker ($M = 29.96$, $SD = 2.30$) than an episode attributed to a fast speaker ($M = 28.98$, $SD = 2.04$), $F(1, 15) = 11.62$, $p = .004$, $\eta^2_p = .44$, whereas no difference in reading time was found between the fast ($M = 29.68$, $SD = 3.19$) and slow speakers ($M = 29.79$, $SD = 3.31$) in the group of participants who did not mimic the speakers, $F(1, 42) = 0.19$, $p = .66$ (see Figure 2).

Moreover, a three-way interaction including affiliation, speaker's gender, and speaking rate was significant, $F(1, 57) = 5.21$, $p = .03$, $\eta^2_p = .08$. When participants were exposed to male speakers' conversations, their reading times differed as a function of speaking rate only in the control condition, $F(1, 57) = 7.36$, $p = .009$, $\eta^2_p = .11$. On the other hand, neither an affiliation and speaking rate interaction ($F(1, 57) = 3.51$, $p = .07$) nor a speaker's sex and speaking rate interaction ($F(1, 57) = 0.28$, $p = .60$) was significant.

As expected, speaker-specific auditory images based on speaking rates influenced the subsequent reading task. Specifically, the influence was found only in participants who consciously mimicked the speakers. In contrast, the effect of the affiliation manipulation was limited. In fact, the speakers were evaluated equally in terms of their personality across the affiliation conditions (control: $M = 3.91$, $SD = 0.60$; face: $M = 4.01$, $SD = 0.48$), $t(63) = 0.76$, $p = .45$. This suggests that the manipulation did not function as a facilitator for enhancing one's mimicry of the speakers.

Study 2

Exposure to schematic faces in Study 1 did not influence participants' evaluations of the speakers and their reading times in the subsequent reading task because the faces, which were not directly related to the speakers' identity and characters, were too subtle as cues to foster participants' mimicry of the speakers. In Study 2, we manipulated the content of the conversation. In addition to the conversation between two speakers used in Study 1, we prepared another conversation by the same speakers, whose content was so technical that undergraduates would likely feel distant from the speakers. By manipulating content of the conversation to lead participants to feel different levels of closeness to the speakers, we expected the technical conversation to decrease participants' level of mimicry of the speakers and accordingly reduce the influence of speaker-specific auditory images on reading. We tested this and the individual differences in conscious mimicry examined in Study 1.

Method

Participants. Sixty-six Japanese undergraduates (27 men and 39 women, $M_{age} = 20.0$ years, $SD = 1.32$) participated in this study. All participants were native Japanese speakers. They were tested individually and paid 500 yen (about 4 USD).

Materials. We prepared familiar and technical conversations. The familiar conversations were identical to those used in Study 1. Technical conversations on a biological experiment were developed in the same manner as the familiar conversations were in Study 1. Each of the technical conversations lasted approximately 90 seconds. The mean speaking rate was 9.47 Japanese syllabary characters per one second among the faster speakers and 7.30 Japanese syllabary characters per one second among the slower speakers. The length and the speaking rates were almost identical to those in the familiar conversations. The speakers were identical to those of the familiar conversation.

Procedure. The procedure was identical to that in Study 1 except for three points. First, each participant was escorted to an experimental booth and seated in front of a desk where only a microphone was set up, as the desktop wallpaper on a computer was not used in this study. Second, participants were assigned to one of the eight conversations (= 2 [conversation] x 2 [speakers' gender] × 2 [speaking rate]). Accordingly, 34 participants were assigned to the familiar conversation condition, whereas the remaining 32 participants were assigned to the technical conversation condition. In addition, there was no bias in the number of male and female participants between male and female conversations (14 male and 20 female participants for male conversation, 13 male and 19 female participants for female conversation). Third, we did not ask participants to evaluate the speakers' personality in a post-questionnaire.

Results and Discussion

Participants could understand the content of the conversation more accurately in the familiar conversation condition (M = 83.1%, SD = 1.04) than in the technical conversation condition (M = 69.9%, SD = 1.39), $t(64) = 3.50$, $p < .001$, suggesting that the content was successfully manipulated as intended.

As in Study 1, we divided participants who reported that they consciously mimicked the speakers from those who did not. Of the 66 participants, 17 of them (10 in the familiar conversation condition and 7 in the technical conversation condition) reported that they consciously mimicked the speakers; the remaining 49 participants (24 in the familiar conversation condition and 25 in the technical conversation condition) did not engage in mimicry. The ratio did not differ across the conversation conditions, $\chi^2(1) = 0.17$, $p = .68$.

Relevant mean reading times are shown in Table 2. All data analyzed in this study are included as a supplementary information file. A 2 (conversation) x 2 (speaker's gender) x 2 (mimicry) x 2 (speaking rate) ANOVA showed that the speaking rate had a significant main effect, $F(1, 58) = 31.74$, $p < .001$, $\eta^2_p = .35$. Similar to the finding of Study 1, participants took a longer time to read an episode attributed to a slow speaker ($M = 30.55$, $SD = 2.96$) than an episode attributed to a fast speaker ($M = 29.98$, $SD = 2.76$). The interaction between mimicry and speaking rate was also significant, $F(1, 58) = 26.42$, $p < .001$, $\eta^2_p = .31$. Consistent with the findings of Study 1, participants consciously mimicking the speakers took a longer time to read an episode attributed to a slow speaker ($M = 31.44$, $SD = 1.85$) than an episode attributed to a fast speaker ($M = 29.38$, $SD = 2.18$), $F(1, 13) = 21.79$, $p < .001$, $\eta^2_p = .63$, whereas no difference in reading time was found between the fast ($M = 30.19$, $SD = 2.92$) and slow speakers ($M = 30.24$, $SD = 3.21$) in the group of participants who did not mimic the speakers, $F(1, 45) = 0.34$, $p = .57$ (see Figure 3).[1]

Moreover, the main effect of the conversation condition was also significant, $F(1, 58) = 5.57$, $p = .02$, $\eta^2_p = .09$. Reading time was longer in the familiar conversation condition ($M = 31.01$, $SD = 2.44$) than in the technical conversation condition ($M = 29.48$, $SD = 3.08$). Although we do not have an obvious interpretation of this main effect, if the familiar conversation is

perceived as less persuasive than the technical conversation, a tendency that slower talking is judged as less truthful and persuasive (Apple, Streeter, & Krauss, 1979) may be reflected in the result. For the purpose of the present research, it is important to focus on the interaction between conversation and speaking rate. Although differences in reading time corresponding to the speakers' speaking rates were greater in the familiar conversation condition (fast: $M = 30.60$, $SD = 2.33$; slow: $M = 31.41$, $SD = 2.50$) than in the technical conversation condition (fast: $M = 29.33$, $SD = 3.05$; slow: $M = 29.63$, $SD = 3.16$), the interaction between conversation and speaking rate was not significant, $F(1, 58) = 0.30$, $p = .58$. Any interaction including conversation condition was not significant, Fs < 3.14, ps $> .08$.

General Discussion

In the two studies, we consistently found that speaker-specific auditory images had an influence on reading. The participants read an episode more slowly when it was attributed to a slow speaker than when it was attributed to a fast speaker. This result is in line with the previous findings of Kosslyn and Matt (1977) and Alexander and Nygaard (2008). Furthermore, we demonstrated that individual differences in one's conscious mimicry of the speaker moderated the influence of speaker-specific auditory images on reading. The influence was confirmed only for those who consciously mimicked the speaker. In contrast, situational manipulations by schematic faces (Study 1) and the content of the speakers' conversation (Study 2) did not moderate the influence of speaker-specific auditory images in reading.

The current findings indicate that Japanese participants could access phonological representations retained from speech during reading, suggesting that the influence of speaker-specific auditory images on reading can be generalized to non-English speakers in non-Western cultural contexts. We expected this based on East Asians' higher sensitivity to

paralinguistic cues such as vocal tone (e.g., Ishii et al., 2003) and their emphasis in focus on the other (Wu & Keysar, 2007). Although the current findings are consistent with this prediction, these tendencies of East Asians also remind us of the importance of cross-cultural testing to see whether East Asians are more likely than Westerners to preserve phonological representations and access them while reading a text. Future cross-cultural investigations will be needed to address the expected differences.

On the other hand, the current findings also suggest that individual differences in one's conscious mimicry of the speaker should be considered as a moderator. Alexander and Nygaard (2008, Study 3) reported that only people with high imagery ability were influenced by the speakers' speaking rates when the content of the reading episode was easy, whereas the influence of auditory images on reading appeared regardless of the imagery ability level when the episode was difficult. Given that the episodes used in Study 1 describe an everyday event among typical undergraduates, which are easy to follow, the current findings corroborate those in the easy text condition of Alexander and Nygaard (2008, Study 3). However, the effect of the individual difference in one's mimicry of the speaker was also found in the technical conversation condition of Study 2, which seems to be inconsistent with the findings of Alexander and Nygaard. As participants' imagery ability was not assessed in the present research, it is impossible to compare the effects of imagery ability and conscious mimicry of the speakers. Despite this, the difference between Alexander and Nygaard (2008, Study 3) and the present research implies that as the technical conversation in this research was not a sufficiently difficult one, the effect of the content of episode (e.g., participants may listen more carefully to a difficult episode than an easy episode, and more careful listening may lead participants to access more phonological representations) might be weak. Instead, the strategical difference in one's

engaging in the reading task (i.e., whether or not participants consciously mimic the speakers) might give rise to a more significant effect on speaker-specific auditory images in reading.

Although we considered situational cues facilitating (or inhibiting) one's mimicry of the speaker as another factor moderating the influence of auditory images on reading, we did not find any effect of either schematic faces (Study 1) or technical conversations (Study 2) as a moderator. We developed and used these stimuli based on previous findings indicating the existence of situational cues enhancing affiliations and interpersonal orientations and closeness to the speakers, which are associated with mimicry. However, in Study 1, schematic faces did not adequately enhance positive evaluations for the speaker, suggesting that neither one's attention to the speakers nor the accompanied mimicry of the speakers was activated. In Study 2, familiar conversations were understood more accurately than technical conversations, suggesting that familiar conversations were easier to understand than technical conversations. However, it was unclear whether familiar conversations perceived as more understandable led participants to feel closeness to the speakers and engage in the mimicry of the speaker. The null effect suggests that our manipulation was so subtle that it was not enough to manipulate the level of affiliations and interpersonal orientations and the level of closeness to the speakers. Thus, the expected situational effects on speaker-specific auditory images should be tested in a more elaborate manner. As in the study by Kurby et al. (2009), providing information on the speakers' characteristics and directly manipulating familiarity with the speakers might be useful in future examinations.

In addition to the problem of our manipulation using subtle situational cues, this study has some limitations that should be considered in future research. First, as mentioned above, we did not assess participants' imagery ability in this research. Thus, it is unclear whether

individual differences in imagery ability influence speaker-specific auditory images in Japanese participants and whether or to what extent one's imagery ability is associated with one's strategy of reading a text (e.g., conscious mimicry of the speakers). Second, although in both studies we found the effect of conscious mimicry of speakers on reading times, the measurement of mimicry was based on one's post-hoc subjective report. By manipulating the level of mimicry of the speakers, future work should add findings supporting the effect of conscious mimicry. Given the validity of auditory imagery (e.g., Stites, Luke, & Christianson, 2013; Yao & Scheepers, 2011), future study conducted in a more rigid manner would present a consistent effect of speaker-specific auditory images in reading regardless of such a manipulation, and experimentally induced mimicry of the speakers would also facilitate the effect. Third, because we divided participants into two groups based on their post-hoc self-reports on their level of mimicry, another unidentified factor characterizing the two groups might have accidentally influenced the results. Finally, while the present research measured reading times, revising the measurement and using a more objective measurement (e.g., eye tracking and brain imaging, see Yao, Belin, & Scheepers, 2011) may be useful for future research because such approaches may be more sensitive to individual differences and potentially cultural differences in the effect of speaker-specific auditory images in reading.

In spite of these limitations, the present research has provided additional evidence on speaker-specific auditory imagery on reading by testing it with non-English speakers in non-Western cultural contexts. Given that data sampling with a more diverse range of cultural backgrounds is needed to generalize psychological findings (Henrich, Heine, & Norenzayan, 2010), it is important to continue further investigations in a more elaborate way for people in non-Western cultural contexts.

Compliance with Ethical Standards

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

References

Alexander, J. D., & Nygaard, L. C. (2008). Reading voices and hearing text: Talker-specific auditory imagery in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 446-459.

Abramson, M., & Goldinger, S. D. (1997). What the reader's eye tells the mind's ear: Silent reading activates inner speech. *Perception & Psychophysics*, *59*, 1059-1068.

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology, 37*, 715-727.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-660.

Gray, H. M., Ishii, K., & Ambady, N. (2011). Misery loves company: When sadness increases the desire for social connectedness. *Personality and Social Psychology Bulletin, 37,* 1438-1448.

Hall, E. T. (1976). *Beyond culture*. NY: Doubleday.

Halpern, A. R. (1988). Mental scanning in auditory imagery for songs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 434-443.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61-135.

Hubbard, T. L. (2010). Auditory imagery: empirical findings. *Psychological Bulletin*, *136*, 302-329.

Hubbard, T. L., & Stoeckig, K. (1988). Musical imagery: Generation of tones and chords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 656-667.

Intons-Peterson, M. J. (1980). The role of loudness in auditory imagery. *Memory and Cognition*,

8, 385-393.

Intons-Peterson, M. J., Russell, W., & Dressel, S. (1992). The role of pitch in auditory imagery. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 233-240.

Ishii, K., Kobayashi, Y., & Kitayama, S. (2010). Interdependence modulates the brain response to word-voice incongruity. *Social Cognitive and Affective Neuroscience, 5,* 307-317.

Ishii, K., Reyes, J. A., & Kitayama, S. (2003). Spontaneous attention to word content versus emotional tone: Differences among three cultures. *Psychological Science, 14,* 39-46.

Ishii, K., & Uchida, Y. (2016). Japanese youth marginalization decreases interdependent orientation. *Journal of Cross-Cultural Psychology*, *47*, 376-384.

Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition and Emotion, 16*, 29-60.

Kosslyn, S. M., & Matt, A. M. (1977). If you speak slowly, do people read your prose slowly? Person-particular speech recoding during reading. *Bulletin of the Psychonomic Society, 9*, 250-252.

Kurby, C. A., Magliano, J. P., & Rapp, D. N. (2009). Those voices in your head: Activation of auditory images during reading. *Cognition*, *112*, 457-461.

Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, *14*, 334-339.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224-253.

Nygaard, L. C., Burt, S. A., & Queen, J. S. (2000). Surface form typicality and asymmetric transfer in episodic memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1228-1244.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 309-328.

Stel, M., & Vonk, R. (2010). Mimicry in social interaction: Benefits for mimickers, mimickees, and their interaction. *British Journal of Psychology*, *101*, 311-323.

Stites, M. C., Luke, S. G., & Christianson, K. (2013). The psychologist said quickly, "Dialogue descriptions modulate reading speed!". *Memory and Cognition*, *41*, 137-151.

Stuart, G. P., & Jones, D. M. (1996). From auditory image to auditory percept: Facilitation through common processes?. *Memory and Cognition*, *24*, 296-304.

Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., & de Gelder, B. (2010). I feel your voice: Cultural differences in the multisensory perception of emotion. *Psychological Science*, *21*, 1259-1262.

van Baaren, R. B., Horgan, T. G., Chartrand, T. L., & Dijkmans, M. (2004). The forest, the trees, and the chameleon: Context dependence and mimicry. *Journal of Personality and Social Psychology*, *86*, 453-459.

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science*, *18*, 600-606.

Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience*, *23*, 3146-3152.

Yao, B., & Scheepers, C. (2011). Contextual modulation of reading rate for direct versus indirect speech quotations. *Cognition*, *121*, 447-453.

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the

mind's ear: A PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience*, *8*, 29-46.

Footnote

1. The mimicry and speaking rate interaction was qualified with speaker's gender, $F(1, 58) = 8.45$, $p = .005$, $\eta^2_p = .13$. As mentioned, the influence of the speakers' speaking rates was found only in participants consciously mimicking. Whereas the pattern appeared in male conversations, $F(1, 58) = 26.29$, $p < .001$, $\eta^2_p = .31$, it became weak in female conversations, $F(1, 58) = 3.25$, $p = .08$, $\eta^2_p = .05$.

Table 1. Mean reading times (second) in Study 1.

	Face		Control	
	M	*SD*	*M*	*SD*
Mimicry				
Female conversations				
Fast speaker	28.31	1.63	29.07	2.77
Slow speaker	29.45	1.12	29.81	3.42
Male conversations				
Fast speaker	30.43	1.43	28.19	1.39
Slow speaker	30.04	2.68	31.22	1.09
No mimicry				
Female conversations				
Fast speaker	29.08	2.27	29.50	2.74
Slow speaker	29.12	2.64	29.60	3.05
Male conversations				
Fast speaker	30.85	3.93	29.22	3.46
Slow speaker	30.93	4.54	29.40	2.71

Table 2. Mean reading times (second) in Study 2.

	Familiar conversation		Technical conversation	
	M	*SD*	*M*	*SD*
Mimicry				
Female conversations				
Fast speaker	29.61	1.74	29.18	1.75
Slow speaker	31.38	1.57	30.51	1.45
Male conversations				
Fast speaker	30.90	1.50	26.13	3.34
Slow speaker	33.05	0.83	30.77	4.02
No mimicry				
Female conversations				
Fast speaker	31.03	2.56	28.30	1.45
Slow speaker	31.73	3.46	28.63	1.77
Male conversations				
Fast speaker	30.61	2.65	30.49	3.73
Slow speaker	30.66	2.11	29.85	4.13

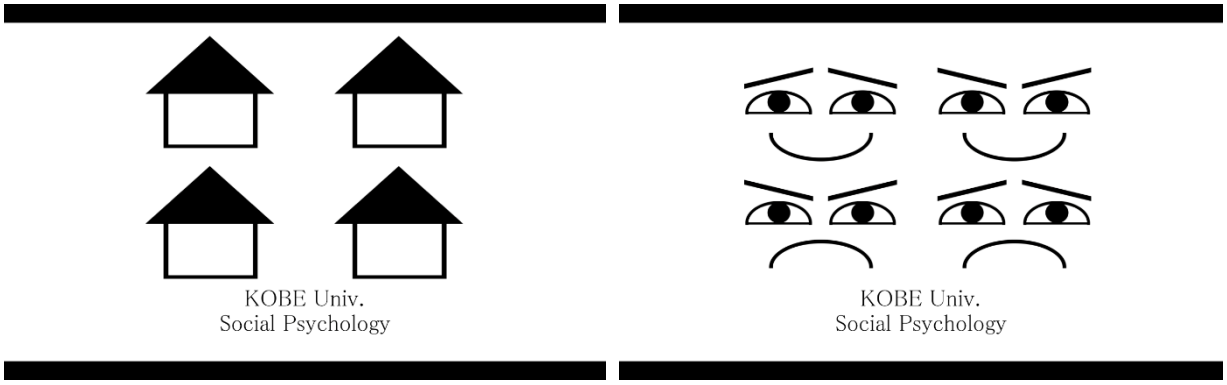

Figure 1. Desktop wallpapers used in Study 1 (left: the control condition, right: the face condition).

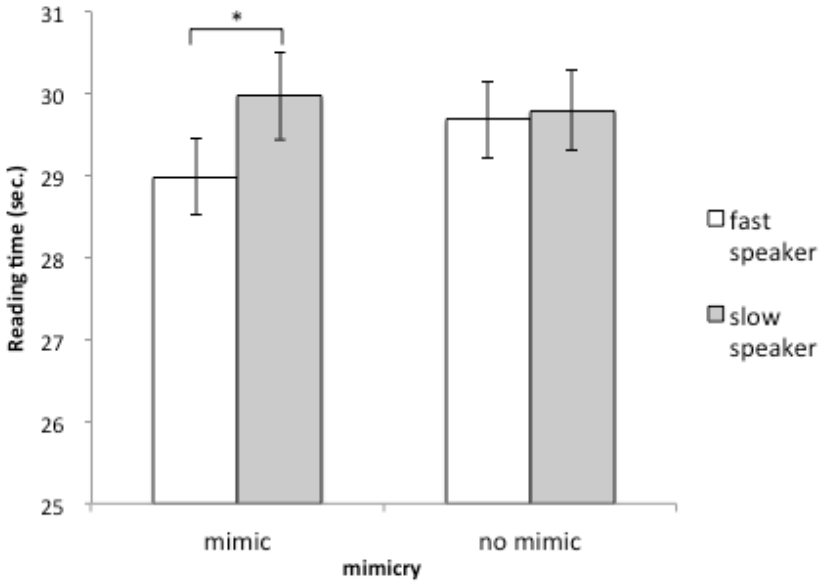

Figure 2. Mean reading times (second) on episodes written by speakers in participants in the two groups of the participants in terms of mimicry in Study 1 (* p = .004)

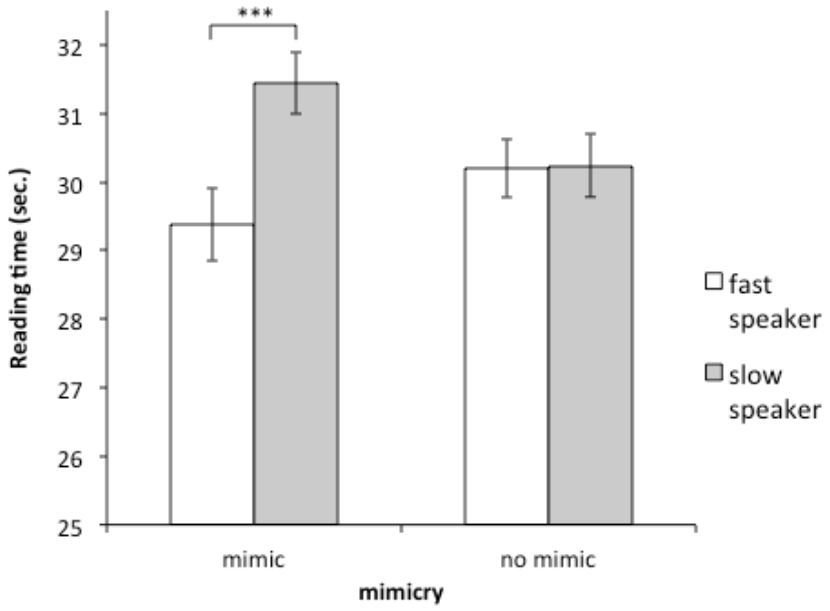

Figure 3. Mean reading times (second) on episodes written by speakers in participants in the two groups of the participants in terms of mimicry in Study 2 ($^{***}p < .001$)