REVERSE LINGUISTIC STEREOTYPING IN ON-LINE PROCESSING: WORD RECOGNITION OF MINIMAL PITCH-ACCENT PAIRS IN TOKYO JAPANESE

Hitoshi Nishizawa, University of Hawai‘i at Mānoa

The study investigates the effects of reverse linguistic stereotyping on on-line spoken word recognition in Tokyo Japanese. Pictures of minimal pitch accent pairs (e.g., /kaki/$_{H*L}$ ‘oyster’ and /kaki/$_{LH}$ ‘persimmon’) were placed at the corner of the screen in addition to two distractors (Visual World Paradigm). The belief about the talker’s ethnicity was manipulated via a talker image in the center of the screen. Native listeners of Japanese (n = 36) were assigned to either Japanese-looking condition or non-Japanese-looking condition. Talkers included two native talkers and two non-native talkers of Japanese for filler trials and one bilingual talker for critical trials. The results showed no effect of face image. The findings contrast with previous works which exclusively employed off-line tasks. This suggests that reverse linguistic stereotyping may not influence on-line word recognition, but rather later stages of language comprehension.

Cite as: Nishizawa, H. (2022). Reverse linguistic stereotyping in on-line processing: word recognition of minimal pitch-accent pairs in Tokyo Japanese. In J. Levis & A. Guskaroska (eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference*, held June 2021 virtually at Brock University, St. Catharines, ON. https://doi.org/10.31274/psllt.13349

INTRODUCTION

Speech perception and comprehension are subject to extra-linguistic factors such as the social information of talkers (e.g., age, ethnicity). Studies in second language (L2) perception and sociophonetics have found that talkers’ social information influences accentedness, comprehension, and phonological perception (e.g., Drager, 2010; Kang & Rubin, 2009). However, previous studies typically employed off-line tasks such as forced-choice tasks, cloze test, with limited work utilizing on-line measurements, for instance, Visual World Paradigm (VWP) (Koops, Gentry, & Pantos, 2008). Moreover, this line of research has not examined Asian languages. This phenomenon should be tested in many languages to help us understand the nature of speech perception. The present study fills this gap by employing VWP to investigate how talkers’ social attribution influences on-line spoken word recognition in Tokyo Japanese.

Literature Review

Studies have shown that L2 speech is less intelligible and more accented than native speech (Munro & Derwing, 2020). However, even when speech is free from a foreign accent, listeners may still perceive an illusory non-native accent and understand less due to their bias about talkers. L2 perception studies suggest reverse linguistic stereotyping (RLS), whereby biased beliefs about talkers’ ethnicity may distort comprehension and detect accentedness (Kang & Rubin, 2009; Rubin, 1992). Biased beliefs about talkers are manipulated by face images to index ethnicity. In this paradigm, the same audio is paired with different images. For instance, Kang and Rubin (2009) paired a native accent with either an Asian face or a Caucasian face, finding that listeners

comprehended less and reported stronger accentedness in the Asian condition relative to the Caucasian condition.

This agreement between the studies which utilized actual L2 speech and RLS studies brings to a question the extent to which these studies concur. A recent psycholinguistic study suggests that L2 speech causes more lexical competition (Lev-Ari, Ho, & Keysar, 2018). Thus, it should be tested if mere beliefs about talkers' ethnicity influence on-line word recognition.

However, studies also found mediating factors. Hanulíková (2018) investigated RLS in Dutch and suggested RLS's subjectivity to listeners' multilingual experiences. She paired a Dutch talker with either Moroccan or Dutch face images. The Moroccan is a distinct ethnicity and indexes a non-Western immigrant community who might speak Dutch without a foreign accent. The results showed no effect of face image on comprehension. She attributed the null effect to the multilingual experiences that native Dutch listeners had, which might have reduced prejudice towards Moroccan ethnicity.

The effects of talker image on speech perception are also investigated in sociophonetics. These studies support the exemplar model (EM), which argues that previous experiences with individuals from a specific community (e.g., acoustic property and social information) are stored in mind (Johnson, 2006). Such studies typically examine the effects of talker image on phonological perception. Drager (2010), for example, utilized four talker images differing in age and found that vowel perception was different across the image conditions. Koops et al. (2008) found that talker image influenced on-line word recognition. They presented a talker image in the center of a screen and four words at each corner (VWP with words). They found differences in the proportion of looks to the competitor (an acoustically similar word to the auditory mentioned word) across the talker image conditions. This indicates that listeners are sensitive to talkers' social information even in on-line language processing.

While L2 perception and sociophonetic studies agree that talker images can influence speech perception, some differences exist. One is that while RLS emphasizes the perceived ethnicity, EM suggests that exemplars, which are formed through previous contacts, are activated by image and influence speech perception. In other words, EM assumes listeners' previous extensive exposure to a specific accent more than RLS does.

The Study

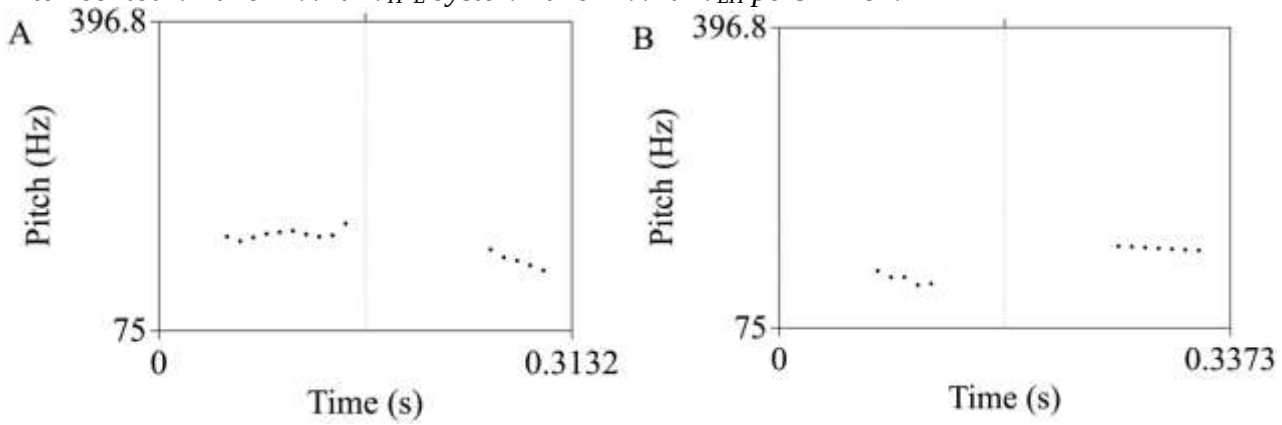
In spite of the difference in the underlying theory, L2 speech perception and sociophonetic studies suggest that talker image influences speech perception. Koops et al.'s (2008) findings further provide insight into on-line language processing. However, RLS has not been tested in on-line measures. Understanding RLS influences on on-line language processing provides insights into at which stage biased belief influences language comprehension.

The present study fills this gap by using VWP to investigate RLS in on-line spoken word recognition in Tokyo Japanese. Tokyo Japanese employs pitch accent (PA) which is realized by a sudden fall in F0 (Kawahara, 2015). It constrains lexical access such that minimal pitch accent pairs (MPAPs) do

not activate each other (Cutler & Otake, 1999). For instance, /kaki/$_{H*L}$ means 'oyster', while /kaki/$_{LH}$ means 'persimmon'. Figure 1 illustrates pitch realization by a male native speaker of Japanese (NSJ).

Figure 1

Pitch contour. Panel A: /kaki/$_{H*L}$ oyster. Panel B: /kaki/$_{LH}$ persimmon.



For learners, acquisition of PA is one of the most challenging phonological aspects and it is strongly associated with comprehensibility (Saito & Akiyama, 2017). That is, PA is closely related to nativeness. Thus, it was posited that perception of PA may be subject to talker image.

Research Questions

> **RQ1** To what extent does the belief about the talkers' non/native status affect on-line word recognition of minimal pitch accent pairs?

To answer the question, two behavioral measures were investigated: mouse clicking and eye-gaze. It was predicted that mouse clicking accuracy and reaction times (RT) and proportion of looks to the target are different as a function of talker image because more lexical competition is expected in the L2 context (Lev-Ari et al., 2018).

METHODS

Listeners

NSJ who grew up in Japan were recruited (n = 36, *Mean age* = 32.08, *SD* = 12.48). At the time of the experiment, all the participants were residing in Oahu, which features significant racial diversity. For instance, 25%, 23.7%, and 22.1% of the residents identified themselves as Filipino, multi-racial, and Japanese, respectively (State of Hawaii, 2018). Because of this multiculturality, all the participants self-reported familiarity with foreign-accented Japanese (*M* = 4.19, *SD* = 1.01 in a scale of five).

Materials

The study had talker image as a between-subject variable and sentential context as a within-subject variable. However, the analysis of the sentential context is not reported in this paper. The study used VWP with 32 target trials (8 pairs * 2 nouns * 2 sentences) and 50 filler trials. Among five male

talkers used, one of them was a bilingual talker. He was used for the target trials by pairing a different talker image across the talker-image conditions. The other talkers were used for filler trials and remained consistent.

Linguistic stimuli

Table 1 shows eight target MPAPs taken from Kindaichi and Akinaga's (2014) dictionary. The MPAPs were embedded in a sentence as illustrated in (1). Filler sentences had the same sentence structure but did not use the target MPAPs. Few filler trials used MPAPs which were not part of the target trials.

(1) サクラの弟はカキを開いた

Sakura	no	otouto	wa	kaki (H*L)	o	hiraita
Sakura	GEN	little bother	TOP	oyster	ACC	open-PST

'Sakura's little brother opened an/the oyster.'

Table 1

Minimal Pitch Accent Pairs

Minimal pitch accent pair	Pitch accent	Meaning	Orthography
Ame	H*L	Rain	雨
Ame	LH	Candy	飴
Botan	H*LL	Peony	牡丹
Botan	LHH	Button	ボタン
Hashi	H*L	Chopsticks	箸
Hashi	LH*	Bridge	橋
Ishi	H*L	doctor	医師
Ishi	LH*	stone	石
Kaki	H*L	oyster	カキ
Kaki	LH	persimmon	柿
kare-	H*LL	flounder	カレイ
kare-	LHH	curry	カレー
Sake	H*L	salmon	鮭
Sake	LH	alcohol	酒
Tsuna	H*L	tuna	ツナ
Tsuna	LH*	rope	綱

Note: * indicates pitch accent.

Speakers

The sentences were recorded by five male talkers: two NSJ, one bilingual (Japanese and English), who left Japan at the age of 18, and two non-native speakers of Japanese (NNSJ). The NSJ and the bilingual talker are from the region where Tokyo Japanese is spoken (Kindaichi & Akinaga, 2014). NNSJ self-reported their overall language proficiency as four on a scale of five (1: elementary, 5: nativelike). The talkers' mean age was 29.2 (*SD* = 2.49). The audio was recorded in a sound-proof booth at a sampling rate of 96kHz.

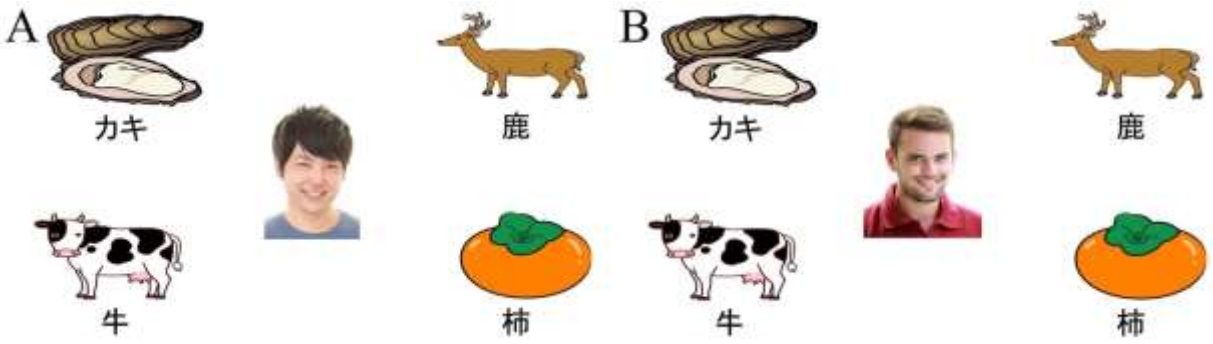
Additional 21 NSJ rated accentedness of talkers using filler sentences (n = 2). The results showed that the bilingual talker was significantly different from the other talkers ($p < .001$). While the two NNSJ were perceived to be more accented than the bilingual talker, he was perceived to be more accented than NSJ (see Appendix A). Despite his ambiguous accentedness, the acoustic analysis suggested accurate production of PA.

Visual stimuli

The visual scene included a talker image, and a picture and orthography of four linguistic objects (target, competitor, and two distractors). The talker image was placed in the center of the screen, while objects were presented at each corner. The talker images should be in a similar age range and should have white background, facial expression, and outfit. For filler trials, two Japanese-looking images and two non-Japanese-looking images were paired with NSJ and NNSJ, respectively. The bilingual talker was paired with either a Japanese-looking image or non-Japanese-looking image (see Figure 2) because his ambiguous accent was most likely to be influenced by talker images (Zheng & Samuel, 2017).

Figure 2

Example displays. Panel A: Japanese-looking condition. Panel B: non-Japanese-looking condition.



Note: カキ /kaki/$_{H*L}$ 'oyster', 鹿 /shika/ 'deer', 牛 /ushi/ 'cow', 柿 /kaki/$_{LH}$ 'persimmon'.

In selecting words for objects, the most frequent orthography (kanji, katakana, or hiragana) was chosen from the Balanced Corpus of Contemporary Written Japanese (Maekawa et al., 2014).

Distractors had an identical number of morae as MPAPs but were phonetically and semantically distinct (see Figure 2). The same distractors were used for a total of four presentations of each MPAP (2 nouns * 2 sentences). Objects did not appear in the same place twice within the target trials.

Other materials

A survey asked the listeners about their hometown and experiences with foreign accents in Japanese.

Procedures

Listeners were randomly assigned to either of the two talker-image conditions. On each trial, a talker image was presented for 1500ms. Subsequently, four objects appeared at each corner, and audio was played after 2000ms of silence. 1500ms after the audio, the next trial was automatically played. The listeners clicked on the object audibly mentioned. The order of the trials was fixed and there was at least one filler trial between target trials. The experiment was individually conducted using an SMI RED250 eye-tracker sampling at 60 Hz in a quiet lab.

Analysis

Before data analysis, some data were excluded; three listeners were not Tokyo Japanese speakers; one listener failed to follow the instructions; five participants had previous contact with one of the talkers; two participants' eye movement was not recognized by the eye-tracker. These two participants with eye recognition issues were included in the mouse click analysis. Thus, the mouse click analysis included 27 listeners and the eye-gaze analysis had 25 (11 for the Japanese-looking condition and 14 for the non-Japanese-looking condition).

Analysis was conducted for the target trials only (i.e., bilingual talker). For the mouse click analysis, accuracy rates RT were examined. The RT analysis excluded incorrect mouse clicks. Accuracy was tested by a generalized linear mixed-effect model (GLMEM), whereas a linear mixed-effect model (LMEM) was used for RT. For the eye-gaze analysis, the target advantage (TA) score was calculated by subtracting the number of 100ms time bins with looks to the competitor from those with looks to the target in the time window from 200ms after the noun onset (Matin, Shao, & Boff, 1993) to the verb onset before mouse clicking. More positive values indicate more looks to the target. The TA score of zero was removed from the analysis (15.82%) as it suggests equal attention to the target and competitor, and the removal resulted in better residual variances. LMEM was also used for the eye-gaze analysis. The effect size was interpreted separately for fixed effects and random effects (Nakagawa & Schielzeth, 2013).

RESULTS

The research question asked the extent to which the belief about the talkers' non/native status affects on-line word recognition. The hypothesis was that mouse click was less accurate and slower, and the proportion of looks to target is less in the non-Japanese-looking condition than in the Japanese-looking condition. The mean mouse click accuracy rate was 96% (*SD* = .19; Japanese-looking condition) and 93% (*SD* = .26; non-Japanese-looking condition). Table 3 shows the GLMEM output

with fixed effects of talker image (non-Japanese-looking condition as intercept) and random intercept by participant and noun. No significant effect was found B = 0.91, *z* = 1.5, *p* = 0.134, CI[-0.25, 2.33]). Only 3% was explained by the fixed effect, while the random effects explained an additional 42% of the variance.

Table 3

The output of generalized linear mixed effect model on mouse click accuracy

					95% CI	
	Estimate	Std. Error	*z*	*p*	*Lower*	*Upper*
Intercept	3.36	0.56	6.01	0.00	2.35	4.65
Talker image	0.91	0.60	1.50	0.134	-0.25	2.33

Mean mouse click RT was 1967 (*SD* = 456) for the Japanese-looking condition and 2079 (*SD* = 439) for the non-Japanese-looking condition. Table 4 illustrates the output of LMEM with fixed effects of talker image (non-Japanese-looking condition as intercept) and random intercept by participant and noun. Echoing the previous model, talker image was insignificant (B – 116.36, t = -1.11, *p* = 0.276, CI[-321.43, 88.63]). Only 2% is explained by the fixed effect, while an additional 37% is explained by the random effects.

Table 4

The output of linear mixed effect model on mouse click reaction times

						95% CI	
	Estimate	Std. Error	*df*	*t*	*p*	*Lower*	*Upper*
(Intercept)	2094.15	85.12	30.49	24.60	0.00	1927.67	2260.74
Talker image	-116.36	104.41	25.13	-1.11	0.276	-321.43	88.63

Figure 3 shows the mean proportion of fixations to different areas of interest from the noun onset to the verb onset. Legend shapes indicate objects, while line type indicates face condition. The mean verb onset was 903ms from the noun onset (*SD* = 106). Talker images received the most eye-gaze at the beginning. The non-Japanese-looking face generally received more looks than the Japanese-looking face. Approximately after 300ms, looks to the target and competitor increased with more looks being on the target. While looks to the target were almost identical across the conditions, the competitor received more looks in the Japanese-looking condition than in the non-Japanese-looking condition, which contradicts the prediction. This is also reflected in the lower mean TA score (2.51, *SD* = 14.4) in the Japanese-looking condition than in the non-Japanese-looking condition (4.03, *SD* = 14.5). Table 5 shows LMEM with fixed effects of talker image (non-Japanese-looking condition as intercept) and random intercept by participant and noun. Talker image was not significant (B = -2.17, *t* = -1.33, *p* > .05, CI[-5.38, 1.05]), indicating that relative looks to the target versus the competitor did not differ between the talker-image conditions. Only .5% is explained by the fixed effect and an additional 4.5% is explained by the random effects.

Figure 3

Mean proportion of fixations to different areas of interest from the noun onset to verb onset

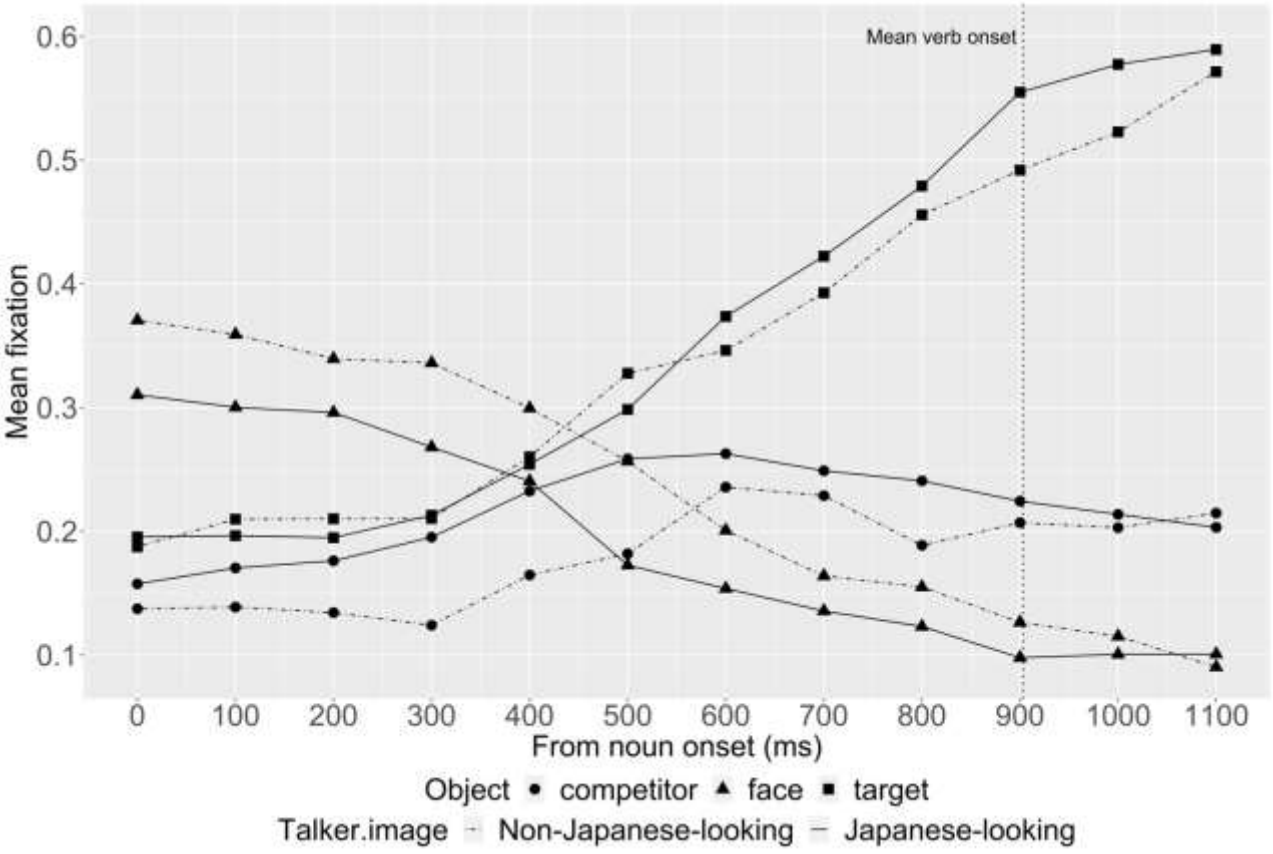


Table 5

The output of linear mixed effect model on eye-gaze

						95% CI	
	Estimate	Std. Error	*df*	*t*	*p*	*Lower*	*Upper*
(Intercept)	4.95	1.39	25.67	3.56	0.001	2.22	7.65
Talker image	-2.17	1.64	22.63	-1.33	0.20	-5.38	1.05

DISCUSSION

The study investigated the extent to which RLS influences spoken word recognition. Neither the mouse-click analysis nor the eye-gaze analysis found the effects. These results contradict the prediction, which suggested that the non-Japanese-looking face causes more lexical competition.

The use of on-line measures was motivated by Koops et al. (2008). However, the absence of effects in the current study may suggest divergences between RLS and EM. As reviewed, RLS does not assume stable exemplars as much as EM does. Therefore, in the present study, listeners might have not had a solid idea of how MPAPs are realized by NNSJ. While the multilingual environment of the research site might have provided some exposure to L2 Japanese, exposure might not be enough to develop solid exemplars.

Yet, few methodological differences from Koops et al. (2008) should be noted. For instance, in their study, the time window started from the onset of a vowel, which differs from a common practice in VWP. The present study followed the common threshold, and the time window started 200ms after the noun onset (Matin et al., 1993). Thus, the different operationalization of the time window might have caused the disagreement. Another difference is the use of pictures for objects; however, it is unlikely the case as pictures should have enhanced the recognition of objects, rather than impeding it.

The multilingual experience might offer another explanation for the null findings. As Hanulíková (2018) suggested, the multilingual and multicultural environment might have mitigated listeners' bias towards a non-Japanese-looking talker. In Hanulíková (2018), comprehension was not influenced by talker image. In the current study, the participants had familiarity with foreign accents in Japanese as evidenced by self-report. Thus, the multicultural experience might have mitigated their bias toward non-Japanese-looking ethnicity, while the exposure might not be sufficient to develop solid exemplars.

Another account is that RLS might not hold in Japanese. Since RLS has been rarely tested in languages other than English, possibilities of some variabilities among languages or cultures cannot be discarded. While Hanulíková (2018) did not attribute the null effects on comprehension to the cross-cultural differences, without sufficient evidence from other contexts and languages, one cannot deny the cross-cultural differences. For instance, Japanese and Dutch are not widely spoken as much as English. To test this hypothesis, one has to control many variables such as the multicultural experience of listeners.

Finally, the null findings suggest that biased belief may not influence on-line spoken word recognition. While previous studies with off-line measures evidence the effects of RLS (Kang & Rubin, 2009), the null finding of the present study suggests that RLS might come into play after word recognition. In fact, this suggestion is fairly in line with Lev-Ari and Keysar (2012), who found that L2 speech is less accurately represented in memory than native speech. They found that native listeners were less accurate in correcting the replaced words after hearing a story in L2 relative to native speech. This finding seems to concur with the RLS studies which found that talker image influenced accuracy in cloze test. Thus, biased belief might influence the representation of speech in memory, but not on-line language processing. This hypothesis seems to be encouraging as reduced comprehension might be due to listeners' unwillingness, which might be alleviated by some intervention (Subtirelu et al., 2022).

There are some limitations to the study. First, larger sample size is required to draw a firmer conclusion. Second, a greater number of sentences for the accentedness rating would have allowed more confidence in interpreting the findings. Third, the bilingual talker's ambiguous accent makes

the direct comparison to the previous studies difficult as previous studies typically used native accents. While Japanese is his L1, his strong L2 proficiency might have influenced his L1 production (Major 1992). That is, he might be perceived as foreign, rather than native in spite of the presentation of the talker image. While this methodology is not entirely new (Zheng & Samuel, 2017), future studies are warranted to address such limitations.

This study concludes by calling for more studies on RLS. Future studies should investigate RLS using a variety of measures, which might include on-line measurements. In addition, RLS should be tested in different contexts and languages to shed more light on the nature of speech perception and possibly language processing.

ACKNOWLEDGMENTS

I would like to thank Dr. Theres Grüter and two anonymous reviewers for their helpful suggestions.

ABOUT THE AUTHOR

Hitoshi Nishizawa is a PhD student at the University of Hawai‘i at Mānoa. He investigates L2 speech from social, psychological and assessment perspectives.

REFERENCES

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of Acoustic Society of America*, *105*, 1877–1888.

Drager, K. (2010). Speaker age and vowel perception. *Language and Speech*, *54*, 99–121.

Hanulíková. A. (2018). The effect of perceived ethnicity on spoken text comprehension under clear and adverse listening conditions. *Linguistics Vanguard*, *4*, 1-9.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*, 485–499.

Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*, *28*, 441-456.

Kawahara, S. (2015). The phonology of Japanese accent, In H. Kubozono (Ed.), *Handbook of Japanese Phonetics and Phonology* (pp. 445-492), Boston, MA: De Gruyter Mouton.

Kindaichi, H. & Akinaga, K. (Eds.). (2014). *Shin Meikai Nihongo akusento jiten (Dai 2-han)* [Shin Meikai dictionary of the Japanese pitch accent (2nd edition)]. Tokyo: Sanseidō.

Koops, C., Gentry, E., & Pantos, A. (2008). The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics: Selected papers from NWAV, 36*, 91–101.

Lev-Ari, S., & Keysar, B. (2012). Less-detailed representation of non-native language: Why non-native speakers' stories seem more vague. *Discourse Processes*, *49*, 523-538.

Lev-Ari, S., Ho, E., & Keysar, B. (2018). The unforeseen consequences of interacting with non-native speakers. *Topics in Cognitive Science*, *10*, 835-849.

Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Koiso, H., Yamaguchi, M., Tanaka, M., & Den, Y. (2014). Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation*, *48*, 345-371.

Major, R. C. (1992). Losing English as a first language. *Modern Language Journal*, *76*(2), 190–208. https://doi.org/10.2307/329772

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception, & Psychophysics*, *53*, 372–380.

Munro, M. J., & Derwing, T. M. (2020). Foreign accent, comprehensibility and intelligibility, redux. *Journal of Second Language Pronunciation*, *6*, 283–309.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133-142.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduate's judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, *33*, 511–531.

Saito, K., & Akiyama, Y. (2017). Linguistic correlates of comprehensibility in second language Japanese speech. *Journal of Second Language Pronunciation*, *3*, 199–217.

State of Hawaii. (2018). Demographic, Social, Economic, and Housing Characteristics for Selected Race Groups in Hawaii.

Subtirelu, N. C., Lindemann, S., Acheson, K., & Campbell, M.-A. (2022). Sharing communicative responsibility: Training US students in cooperative strategies for communicating across linguistic difference. *Multilingua*, *0*.

Zheng, Y., & Samuel, A. G. (2017). Does seeing an Asian face make speech sound more accented? *Attention, Perception, & Psychophysics*, *79*, 1841–1859.

APPENDIX A

The output of linear mixed effect model on accentedness

	Estimate	Std. Error	*t*	*p*	95% CI	
					LL	*UL*
(Intercept)	49.667	5.080	9.777	0.000	40.305	59.028
NNSJ1	45.929	5.352	8.581	0.000	35.499	56.358
NNSJ2	31.929	5.352	5.966	0.000	21.499	42.358

NSJ1	-46.048	5.352	-8.604	0.000	-56.477	-35.618
NS2J	-37.429	5.352	-6.993	0.000	-47.858	-26.999

Note: Higher value means stronger foreign accent (1-100). Bilingual talker as an intercept.