Speech Prosody 2016
31 May - 3 June 2016, Boston, USA

Does speech production in L2 require access to phonological representations?

Yuki Asano[12], *Bettina Braun*[2]

[1] University of Konstanz, Germany
[2]University of Tübingen, Germany
yuki.asano@uni-tuebingen.de, bettina.braun@uni-konstanz.de

Abstract

Following the theory of direct realism [1, 2, 3], non-native (L2) speakers should be able to imitate a stimulus without requiring the access to L2 phonological representations. In line with theories of working memory [4, 5], however, they should encounter difficulties in imitating a stimulus with L2 phonological structure at the point once phonetic information decayed and therefore phonological representations are required. In order to test the validity of these claims, the current study investigates L2 speakers' ability to immitate L2 segmental length contrasts in an immediate vs. delayed imitation paradigm. In the immediate imitation condition, participants should be able to make use of phonetic information taken from the acoustic echo of the stimuli. In the delayed imitation condition, however, phonological representations were required after the decay of phonetic information. The results show that L2 speakers' performance differed from that of native (L1) speakers in the immediate imitation condition, suggesting that phonological representations had been already activated in the immediate condition. L2 speech production may inevitably require phonological representations. The claim made by the direct realist view was not supported in this study.

Index Terms: immediate vs. delayed imitation, L2, consonant length contrasts

1. Introduction

Producing L2 consonant length contrasts has been reported to be difficult for L2 speakers when their L1 does not bear such lexical contrasts [6, 7, 8, 9]. Kabak *et al.* showed that native-like timing of geminate consonants was difficult to produce, even for advanced German learners of Italian who had considerable exposure to Italian [6]. They found significant differences in the geminate-singleton duration ratios of nonsense word minimal pairs across groups (non-learners < advanced learners < Italian L1 speakers). Also, Han showed that the timing control of L2 geminate and single stop consonants was challenging, even for advanced L2 American English learners of Japanese [7]. While L1 speakers distinguished between geminate and single stops by controlling the closure durations in a mean ratio of 2.8:1.0 (ranging from 2.5 to 3.2:1.0), L2 learners pronounced the same tokens in a diverse and random manner (mean = 2.0:1.0, ranging from 0.9 to 4.0:1.0). Such difficulties found in previous production studies are normally explained by the fact that appropriate L2 phonological contrasts were lacking in the L2 speakers' mental representations.

The current study examines whether L2 speakers are able to produce L2 segmental length contrasts in an imitation task when they do not necessarily require phonological representations. Using an immediate vs. delayed imitation paradigm [10], German L2 learners of Japanese, German non-learners and Japanese L1 speakers were tested in their imitation accuracy of both consonant and vowel length contrasts, the latter serving as a reference because only the former are not lexical in German.

The direct realist theory [1, 2, 3] argues for a perspective wherein speech perception and production are closely linked. It claims that a listener directly apprehends the perceptual object and does not solely apprehend representative or abstract features from which the object must be inferred [11]. According to the theory, speech production is driven by reflexive phonetic gestures that are mediated automatically in speech perception without requiring access to the speakers' phonological representations. If this account is valid, the lack of L2 phonological representations should not impede sound imitation with an L2 phonological structure, and L2 speakers' imitation accuracy should not differ from that of L1 speakers in either the immediate or delayed imitation conditions.

Following the theory of working memory [4, 5], the acoustic echo of stimulus words is still available and participants should be able to use it to execute a phonetic plan [12, 13]. In the delayed imitation condition, continuous interactions occur between working memory and long-term memory to maintain and refresh the phonetic information of the stimulus through inner speech or rehearsal the acoustic information while waiting to speak. Each time an echo that is held in working memory is communicated to long-term memory, the feedback loop forces a subsequent echo toward the central tendency of the stored category. In this way, feedback from long-term memory progressively assimilates the L2 stimulus towards the L1 categories if the L2 category is not yet very robust [14, 15]. Idiosyncratic details of the original imitation stimulus will be attenuated in the eventual echo used for output [10]. After several seconds [16, 10], the echo in working-memory will be the L1 mental category prototype and no longer what was initially perceived. Therefore, imitation accuracy of L2 speakers should decline in the delayed imitation.

A frequently cited study implementing this immediate vs. delayed imitation paradigm is Goldinger's work [10]. The duration of the delay was 4000 ms. In the current experiment, however, the duration was set to 2500 ms to keep the experiment short enougth that participants would not lose their concentration or motivation, while at the same time ensuring that processing taps into the phonological level after phonetic information decays (= 2000 ms, [17, 16]). To our knowledge, there are no other theoretical arguments that call for 4000 ms instead of 2500 ms.

In the previously published authors' work [18] , an AX-task was conducted with short and long inter-stimulus intervals (= ISIs, 300 ms vs. 2500 ms), in which the same German participants as in the current imitation study had to discriminate the same Japanese (L2) consonant length contrast stimuli. There-

10.21437/SpeechProsody.2016-49

fore, the duration of the imitation delay in the current study corresponds to the duration of long ISIs in the previous work. This experimental paradigm allowed for comparison of the results of the perception and production experiments while also localising possible sources of foreign accents in L2 perception, mental representations and production.

Further, in order to ensure that L2 speakers could rely solely on the acoustic echo of the stimuli without having any difficulties in articulation, the following aspects were considered: First, the investigation on segmental length contrasts excluded difficulties related to articulatory or motor processes [19, 20, 21, 22]. Languages vary in their phonetic settings [19, 23, 24, 21], i.e. how to configure the vocal apparatus (such as lip, tongue, jaw) for language-specific habits. L2 learners may have difficulties producing L2 prosodic contrasts because they may not be able to coordinate the vocal apparatus in the way required to produce the L2 contrast. However, difficulties relating to the phonetic setting should not be relevant for producing segmental length contrasts, as the realisation of segmental length contrasts does not require different coordination of the articulatory apparatus for short and long segments. Second, the stimuli only contained sounds that are easy to produce for both Japanese and German speakers to exclude difficulties in that area. Using the simplest imitation task, tracing back to the basis of speech production was the goal of the study.

Accuracy of imitation was defined as how similar the productions were with respect to the stimulus words. More precisely, duration ratios of the critical consonants of the stimuli were subtracted from those of the participants' productions. Duration ratios between singleton and geminate consonants have been extensively analysed in previous studies [25, 26, 27], and are claimed to be the major acoustic correlate and perceptual cue for the distinction between single/geminate consonants in Japanese.

2. Experiment

2.1. Methods

2.1.1. Participants

Twenty-four Japanese L1 speakers (10 male, 20-31 years), 24 L1 German speakers who were not learning Japanese (= non-learners, 8 male, 19-30 years) and 48 L2 German learners of Japanese (30 male, 20-34 years) took part in this study after the perception experiments conducted by the author [18]. They were all unaware of the purpose of the experiment, and none of the learners had prior training in Japanese phonology.

2.1.2. Materials

Twenty-one disyllabic triplets of pseudo-words, which differed segmentally only in the length of the first vowel or in the length of the second consonant (e.g., [pɯnɯ],[pɯːnɯ], [pɯn:ɯ]), were created. Prior to the experiment, they were evaluated in a pretest with Japanese and German L1 listeners (different from those of the main experiment) to select only stimuli that did not activate a word via phonological analogy. Participants were presented with one stimulus at a time and were required to write down the first word that came to mind. Responses from 24 Japanese and 24 German L1 listeners were analysed separately. The six non-word triplets with the lowest association strengths in both groups were selected (the word association rate for the selected six triplets ranged between 29.8 % and 45.3%, mean = 34.5%, while that of all 21 triplets was between 29.8% and 100.0%, mean = 52.3%). The selected stimuli differed in manner of articulation and voicing of the medial consonant (= phon), *punu, gunu, gupu, gubu, zusu, sufu*. Moreover, stimuli were varied by adding a pitch fall that occurred simultaneously with the short vs. long consonant, resulting in flat vs. falling pitch. Producing geminates in the falling pitch condition was expected to be more difficult than that in the flat pitch condition, because in the former case, speakers had to coordinate both segmental length and pitch simultaneously. The materials were recorded by a female speaker of Japanese in two pitch conditions: high flat pitch and falling pitch (with a pitch fall during the medial consonant pitch tracks, see Figure 1). Note that the same recordings were used in the perception experiment conducted by the author [18]. The average pitch range in the flat pitch condition was 1.3 semitones, ranging between 1.0 and 1.6 semitones while that of the falling pitch condition was 13.0 semitones, ranging between 10.5 and 16.4 semitones.

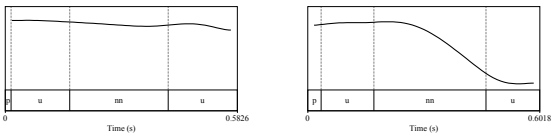

Figure 1: *F_0 track of geminate stimuli (punnu) in the flat and falling pitch conditions. F_0 range is between 100 and 350 Hz.*

To verify the durational differences of the stimuli, the durations of the short and long consonants and vowels were analysed. A linear mixed effects regression (LMER) model with the vowel or consonant *duration* as dependent measure and *pitch* (flat vs. falling) and *segmental length condition* (short vs. long vowel or consonant) as fixed factors and *phon* as a random factor including random slopes for the fixed factors [28, 29] showed a significant interaction between *pitch* and *segmental length condition* for vowel length contrasts ($p < 0.01$) but not for consonant length contrasts ($p > 0.1$). Therefore, as a second step, the durations in the flat and falling pitch conditions were analysed separately for vowel length contrasts. Results of paired t-tests showed that, on average, long vowels in the flat pitch condition were 3.3 times longer than short vowels ($t(5) = 20.0$, $p < 0.001$), and those in the falling pitch condition were 3.0 times longer ($t(5) = 28.1$, $p < 0.001$). For consonant length contrasts, the durations in the flat and falling pitch conditions were analysed together. Paired t-test results showed that geminates were on average 3.2 times longer than singleton consonants ($t(10) = 25$, $p < 0.001$). These duration measurements ensured that the acoustic criteria for the length distinction in vowels and consonants were met (the ratio for Japanese vowels was approximately 3.2:1, [30] and for consonants 3.2:1 [8, 31]).

2.1.3. Procedure

Since the sex of the speaker of recorded experimental stimuli does not seem to influence imitation performance [32], the recordings produced by the female speaker were used for both male and female participants. The female voice was dropped by 4 semitones to roughly match the mean F_0 of the male participants.

One pseudo-randomised experimental list was constructed by presenting all stimulus words (totalling in 36 trials). The same list was used for both the immediate and delayed imitation tasks. In total, each participant imitated 72 words (36 stimuli x 2 imitation conditions) in the given order. The constraints for the randomisation kept at least 3 trials between the stimuli of

the same reference phon (e.g. *punu*, *gunu*) and at least 2 trials between the stimuli of the same segmental length condition. In this way, similar stimuli were presented separately with a certain distance between them. The experiment was programmed and presented using *Presentation* (Neurobehavioral Systems). Auditory stimuli were presented via headphones (Sony MDR-CD570).

All participants took part in the immediate imitation task first and then in the delayed imitation task. Before each task, participants were given a written instruction of the experiment and the procedure. They were instructed to imitate stimuli as correctly as possible after a cross appeared on the screen. The aim of the study was not communicated to the participants. All written instructions were given in English in the same way for all three participants groups. Both the immediate and delayed imitation tasks began with the same 10 training trials that were not used as experimental items. After training, there was a pause (1 minute) before the main experimental task. Each trial began with a sinusoid beep of 44100 Hz (500 ms) followed by 500 ms of silence. After this start signal, the auditory stimulus was presented. In the immediate imitation condition, a cross was shown at the offset of the stimulus. In the delayed imitation condition, it was shown 2500 ms after the offset of the stimulus. Participants were then given a maximum of 2500 ms to imitate before timeout. The intertrial-interval after the timeout was 1000 ms. No feedback was provided during the experiment. Participants' responses were recorded using a portable digital speech recorder (M-Audio Micro Track II Digital-Recorder) via a microphone with a 41kHz sampling rate and 16 bit stereo format.

2.1.4. F_0 extraction and segmental annotation

In total, 6912 data points were recorded (96 participants x 72 trials). Segmental boundary annotation was carried out on the recorded raw data using Praat, applying standard segmentation criteria [33]. Five segmental boundaries were considered: | C| V| C| V| , | C| V| C:| V| and | C| V:| C| V| .

2.2. Results

From the raw durations, normalised relative ratios of short-long consonants were calculated in the following manner: First, the relative durations (with respect to total durations) of a long consonant were divided by those of a short consonant (= ratios of long-short consonant). Then, the ratios of the stimuli were subtracted from those of the participants' productions. In the same way, normalised relative ratios of short-long vowels were calculated. The value 0 indicates a ratio that was the same as that of the stimuli. Positive values indicate that the ratios of the productions were larger than those of the stimuli. There were 4608 ratios in the analysis.

In the following analyses, statistical results from LMER models are reported and descriptive mean values and 95% CI error bars are shown in plots as a complementary visual data analysis [34, 35, 36, 37, 38, 39].

The LMER-analysis on the normalised consonant ratios showed a significant main effect of *language group* (the ratios produced by the learners were smaller than those produced by the Japanese, ß= -0.42, SE = 0.05, t = -8.7, $p < 0.001$, the ratios produced by the non-learners were smaller than those produced by the Japanese, ß= -0.58, SE = 0.06, t = -10.6, $p < 0.001$, the ratios produced by the non-learners were smaller than those produced by the learners, ß= -0.17, SE = 0.05, t = -3.5, $p < 0.001$). Moreover, there was an interaction between *language group* and *pitch* (the difference between the ratios produced by the learners and by the Japanese became smaller in the falling pitch condition when compared to the flat pitch condition, ß= -0.21, SE = 0.04, t = -5.0, $p < 0.001$; and that between the ratios produced by the non-learners and by the Japanese, ß= -0.22, SE = 0.05, t = -4.4, $p < 0.001$; no interaction between the learners and the non-learners, p = 0.9).

The LMER-analysis on the normalised vowel ratios showed a significant main effect of *language group* (the ratios produced by the non-learners were larger than those produced by the Japanese, ß= 0.17, SE = 0.08, t = 2.2, $p < 0.03$; the ratios produced by the learners tended to be larger than those produced by the Japanese, ß= 0.12, SE = 0.06, t = 1.8, p = 0.08; the two German groups did not differ from each other, p = 0.4). Moreover, there was an interaction between *language group* and *pitch* (the difference between the ratios produced by the learners and by the Japanese became much larger in the falling pitch condition than in the flat pitch condition, ß= 0.17, SE = 0.05, t = 3.5, $p < 0.001$; and this was also true for the ratios produced by the non-learners and by the Japanese, ß= 0.13, SE = 0.06, t = 2.3, $p < 0.03$; however, no interaction was found between the learners and non-learners, p = 0.5).

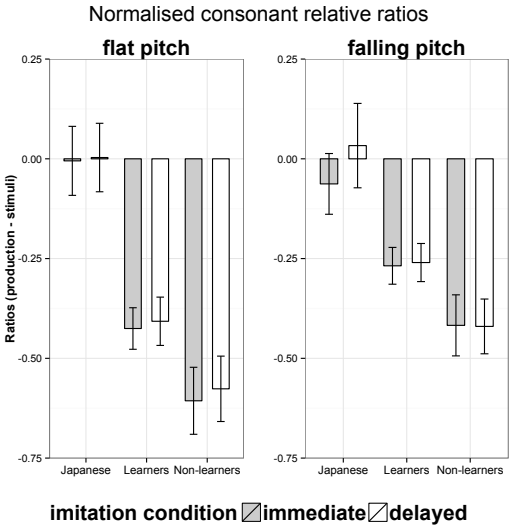

Figure 2: *Mean normalised consonant ratios and 95% CI bars in the flat pitch condition (left) and in the falling pitch condition (right) for each language group and imitation condition.*

Figures 2 and 3 show that the ratios produced by the Japanese speakers did not differ from those of the stimuli in any condition. As for the consonant duration ratios, those produced by the learners were smaller than those of the stimuli, and those produced by the non-learners were much smaller overall. This was true in both pitch conditions. The ratios produced by the learners and the non-learners differed from one another. In the falling pitch condition, the differences between the stimuli and the productions of learners and non-learners became smaller compared to the Japanese group. As for the vowel duration ratios, the ratios produced by the Japanese did not differ from those of the stimuli in the flat or falling pitch condition. The ratios produced by learners and non-learners were larger than those of the stimuli in both pitch conditions. The ratios of the non-learners were larger than those of the learners in the immediate imitation and flat pitch conditions. The two groups did not differ from one another in any other way. In the falling pitch

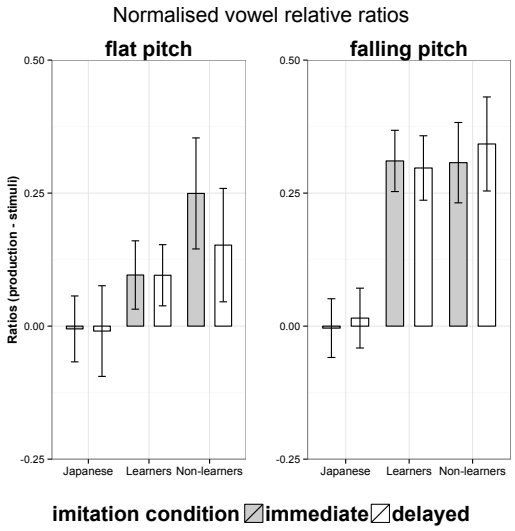

Figure 3: *Mean normalised vowel ratios and 95% CI bars in the flat pitch condition (left) and in the falling pitch condition (right) for each language group and imitation condition.*

condition, differences between the stimuli and the productions of the learners and the non-learners became greater compared to the Japanese group.

3. Discussion

The analysis of duration ratios showed that Japanese speakers produced the same duration ratios as those of the stimuli regardless of the type of contrasts (vowel or consonant duration ratios) and regardless of pitch and imitation conditions, thus acting as a suitable reference group. As for the L2 speakers' productions, the learners' consonant duration ratios were smaller than those of the stimuli and the non-learners' ratios were even smaller. The learners' performance was in this sense better than that of non-learners. This result can be regarded as a positive effect of exposure to the L2. Interestingly, there was no effect of imitation condition, suggesting that it was difficult for the German participants to produce the L2 segmental length contrasts even in the immediate imitation task. Contrary to consonant duration ratios, learners and non-learners produced larger vowel duration ratios than those of the stimuli in both pitch conditions and imitation conditions. The L2 speakers may have exaggerated the contrasts that were familiar in their L1. This was all the more true when non-learners imitated the stimuli in the easiest task condition (i.e. in the immediate imitation and flat pitch conditions).

The data in this study do not support the claim made by the direct realist view that an immediate imitation is driven by reflexive phonetic gestures that are mediated automatically in speech perception without requiring access to the speaker's phonological representations, because performance by the L1 and L2 speakers already differed in the immediate imitation condition. The phonetic details of the stimuli that were still available in the immediate imitation condition were not advantageous to L2 speakers. In order to explain this finding, it is plausible to claim that L1 phonological representations might have been activated already in the immediate imitation condition independently from the necessesity of their activation. This claim is supported by studies that show that auditory information may directly co-activate L1 phonological representations [40, 41]. Another piece of evidence that supports this claim is also found in my previous work [18], in which reaction times in discrimination tasks were longer for learners than non-learners. This difference was observed regardless of ISI condition. This finding can be regarded as evidence that the learners established L2 phonological representations and competed with L1 representations. They thus needed more time to select either L1 or L2 phonological representations, while non-learners did not experience that selection difficulty. Importantly, this happened in the short ISI condition, in which the activation of phonological representations were not required.

In my previous investigation [18], native-like discrimination ability by learners and non-learners was found only in the short ISI and flat pitch conditions, in which the task demands were lowest. Once the ISI became longer, non-learners' discrimination ability decreased and differed from that of L1 listeners. The result in the perception experiment [18] therefore suggests that L2 listeners had difficulties in speech perception once processing required more phonological representations. In both my previous investigation and the current study, Japanese L1 listeners discriminated and imitated segmental length contrasts equally well in all experimental conditions, thus serving as an appropriate reference group. As for German learners' and non-learners' performance, despite their fairly good discrimination ability in the perception experiment, their imitation performance in terms of the duration ratios of vowel and consonant length contrasts differed from the ratios of the stimuli and those by the Japanese L1 speakers in the immediate imitation and flat pitch condition(again when the task demands were lowest). While L2 speakers could discriminate L2 segmental length contrasts under the lowest task demands, they failed to produce the contrasts under the comparable experimental situation (in the immediate imitation and flat pitch conditions). This finding suggests that the acoustic correlates of the contrasts were not sufficient for successful imitation of the contrasts, but speech production may have inevitably required phonological representations. Taken together, the findings suggest that L2 speakers' difficulties in processing of L2 prosodic contrasts relate to their mental representation stage.

4. Conclusion

Using the immediate vs. delayed imitation paradigm, the study investigated whether L2 speakers were able to imitate stimuli that contrasted in segmental length in an immediate and delayed imitation task. The findings show reduced temporal contrasts even in immediate imitation, which suggests that L2 speakers could not directly reproduce the original stimuli. We argue that it is participants stored mental representations of consonant length that affected their L2 imitation success. Together with my our previous work [18], L2 speakers' difficulties in processing L2 segmental length contrasts were shown to relate to their mental representation stage in both L2 perception and production.

5. Acknowledgements

This work was carried out under the project "Perception, storage and articulation of second language phonology" supported by Young Scholar Fund at the University of Konstanz.

6. References

[1] C. A. Fowler, "An event approach to the study of speech perception from a direct-realist perspective," *Journal of Phonetics*, vol. 14, pp. 3–28, 1986.

[2] ——, "Listener-talker attunements in speech," *Haskins Laboratories Status Report on Speech Research*, vol. SR-101/102, pp. 110–129, 1990.

[3] ——, "Sound-producing sources as objects of perception: Rate normalization and nonspeech perception," *Journal of Acoustical Society of America*, vol. 88, pp. 1236–1249, 1990.

[4] A. D. Baddeley and G. J. Hitch, "Working memory," in *The psychology of learning and motivation*, G. H. Bower, Ed. London: Academic Press, 1974, vol. 8, pp. 47–90.

[5] A. Baddeley, S. Gathercole, and C. Papagno, "The phonological loop as a language learning device." *Psychological Review*, vol. 105, no. 1, pp. 158–173, 01 1998.

[6] B. Kabak, T. Reckziegel, and B. Braun, "Timing of second language singletons and geminates," in *The 17th International Congress of Phonetic Sciences in Hong Kong, 17-21 August 2011*, 2011, pp. 994–997.

[7] M. S. Han, "The timing control of geminate and single stop consonants in Japanese: A challenge for nonnative speakers," *Phonetica*, vol. 49, pp. 102–127, 1992.

[8] ——, "Acoustic manifestations of mora timing in Japanese," *Acoustical Society of America*, vol. 96, no. 1, pp. 73–82, 1994.

[9] J. Mah and J. Archibald, "Acquisition of L2 length contrasts," in *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference*, Ottawa, USA, 2003, pp. 208–212.

[10] S. D. Goldinger, "Echoes of echoes? An episodic theory of lexical access," *Psychological Review*, vol. 105, pp. 251–279, 1998.

[11] C. T. Best, "A direct realist view of cross-language speech perception," in *Speech perception and linguistic experience*, W. Strange, Ed. Timonium MD: York Press, 1995.

[12] J. M. Levelt, Willem, *Speaking. From intention to articulation*. Cambridge, MA: MIT Press, 1989.

[13] ——, "Producing spoken language: A blueprint of the speaker," in *The neurocognition of language*, C. Brown and P. Hagoort, Eds. Oxford University Press, 1999, ch. 4, pp. 83–122.

[14] P. K. Kuhl, "Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not," *Perception & Psychophysics*, vol. 50, no. 2, pp. 93–107, 1991.

[15] P. K. Kuhl and P. Iverson, "Linguistic experience and 'perceptualmagnet effect'," in *Speech perception and linguistic experience*, W. Strange, Ed. Timonium MD: York Press, 1995.

[16] A. D. Baddeley, *Working memory*. Oxford: Oxford University Press, 1986.

[17] A. Baddeley, "The episodic buffer: a new component of working memory?" *Trends in Cognitive Sciences*, vol. 4, no. 11, pp. 417–423, 2000.

[18] Y. Asano, "Stability in perceiving non-native segmental length contrasts," in *Proceedings of the 7th International Conference on Speech Prosody*, Dublin, Ireland, 2014, pp. 321–325.

[19] J. H. Esling and R. F. Wong, "Voice quality settings and the teaching of pronunciation," *TESOL Quarterly*, vol. 17, no. 1, pp. 89–95, 03 1983.

[20] J. Kerr, "Articulatory setting and voice production: Issues in accent modification," *Prospect*, vol. 15, no. 2, pp. 4–15, 2000.

[21] I. Mennen, J. Scobbie, E. De Leeuw, F. Schaeffler, and S. Schaeffler, "Measuring language-specific phonetic settings," *Second Language Research*, vol. 26, no. 1, pp. 191–215, 2010.

[22] W. Strange, "Cross-language phonetic similarity of vowels: Theoretical and methodological issues," in *Language Experience in Second Language Speech Learning: in honorof James Emil Flege*, O.-S. Bohn and M. J. Munro, Eds. John Benjamins, 2007, pp. 35–55.

[23] B. Honikman, "Articulatory settings," in *In honour of Daniel Jones*, D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, and J. L. M. Trim, Eds. Longman, 1964, pp. 73–84.

[24] J. Laver, *Principles of phonetics*. Cambridge: Cambridge University Press, 1994.

[25] H. Fujisaki, K. Nakamura, and T. Imoto, "Auditory perception of duration of speech and non-speech stimuli," in *Auditory Analysis and Perception of Speech*, G. Fant and M. A. A. Tatham, Eds. London: Academic Press, 1975, pp. 197–219.

[26] T. Harada, "The acquisition of single and geminate consonants by English-speaking children in a Japanese immersion program," *Studies in Second Language Acquisition*, vol. 28, no. 4, pp. 601–632, 2006.

[27] Y. Hirata and J. Whiton, "Effects of speaking rate on the single/geminate stop distinction in Japanese," *Journal of Acoustical Society of America*, vol. 118, no. 3, pp. 1647–1660, 2005.

[28] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of Memory and Language*, vol. 68, no. 3, pp. 255–278, 2013.

[29] I. Cunnings, "An overview of mixed-effects statistical models for second language researchers," *Second Language Research*, vol. 28, no. 3, pp. 369–382, 2012.

[30] S. Akaba, "An acoustic study of the Japanese short and long vowel distinction," 2008.

[31] Y. Homma, "Durational relationships between Japanese stops and vowels," *Journal of Phonetics*, vol. 9, no. 3, pp. 273–281, 1981.

[32] L. L. Namy, L. C. Nygaard, and D. Sauerteig, "Gender differences in vocal accommodation:," *Journal of Language and Social Psychology*, vol. 21, no. 4, pp. 422–432, 2002.

[33] O. Turk, M. Schöder, B. Bozkurt, and L. Arslan, "Voice quality interpolation for emotional text-to-speech synthesis," in *Proceedings of the 7th Interspeech*, Lisbon, 2005, pp. 797–800.

[34] J. Cohen, "The earth is round ($p < .05$)," *American Psychologist*, vol. 49, pp. 997–1003, 1994.

[35] G. Cumming, *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-analysis*. New York: Routledge, 2011.

[36] ——, "The new statistics: Why and how," *Psychological Science*, 2013.

[37] V. E. Johnson, "Revised standards for statistical evidence," *Proceedings of the National Academy of Sciences*, 2013.

[38] G. R. Loftus, "A picture is worth a thousand p values: On the irrelevance of hypothesis testing in the microcomputer age," *Behavior Research Methods, Instruments, & Computers*, vol. 25, no. 2, pp. 250–256, 1993.

[39] J. P. Simmons, L. D. Nelson, and U. Simonsohn, "False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant," *Psychological Science*, vol. 22, no. 11, pp. 1359–1366, 2011.

[40] I. Darcy, L. Dekydtspotter, R. A. Sprouse, J. Glover, C. Kaden, M. McGuire, and J. H. Scott, "Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English– L2 French acquisition," *Second Language Research*, vol. 28, no. 1, pp. 5–40, 2012.

[41] R. P. Wayland and S. G. Guion, "Training English and Chinese listeners to perceive Thai tones: A preliminary report," *Language Learning*, vol. 54, no. 4, pp. 681–712, 2004.