LANGUAGE LEARNING
A Journal of Research in Language Studies
Language Learning ISSN 0023-8333

EMPIRICAL STUDY

The Efficacy of Gesture on Second Language Pronunciation: An Exploratory Study of Handclapping as a Classroom Instructional Tool

Takehiro Iizuka,[a] Kimi Nakatsukasa,[b] and Aaron Braver[b]
[a]University of Maryland [b]Texas Tech University

In this study, we examined the efficacy of gestures for the acquisition of L2 segmental phonology. Despite teachers' frequent use of gestures in the classroom to teach pronunciation, the field lacks empirical support for this practice. We attempted to fill this gap by investigating the effects of handclapping on the development of L2 Japanese segmentals (long vowels, geminates, and moraic nasals). We assigned L1 English university students in beginning Japanese courses to one of two groups where they practiced pronouncing the targets with or without handclapping in the classroom. They also completed picture elicitation (production) and dictation (perception) tasks as pretests, immediate posttests, and delayed posttests. The results show that, on the delayed perception posttest, only those who saw and performed handclapping maintained the instructional effect, indicating that the memory-enhancing effect of gestures, at least in the form of handclapping, might reach the level of segmental phonology in L2 acquisition.

This article is dedicated to the memory of the second author, Kimi Nakatsukasa, who passed away young during the journal review process. Without her guidance, this paper would have never been published. She was also an exceptional person who lit up the lives of all whom she touched. We would also like to thank the following individuals: Hidetoshi Tanihara, Koyuki Mitani, and Taichi Yamashita for their support with several aspects of this project including material development and scoring and Yukari Hirata and Spencer Kelly for their insightful comments on an earlier version of the manuscript. Last but not least, we would like to express our sincere gratitude to the journal editors and four anonymous reviewers for their very useful comments that helped us shape the final version with much better quality.
The handling editor for this article was Judit Kormos.

Correspondence concerning this article should be addressed to Takehiro Iizuka, Department of Second Language Acquisition, 3215 Jimenez Hall, University of Maryland, College Park, MD 20742, USA. E-mail: tiizuka@terpmail.umd.edu

DOI: 10.1111/lang.12415

Keywords gesture; segmental phonology; pronunciation instruction; productive/receptive L2 knowledge; Japanese; foreign language

Introduction

Pronunciation has long been considered one of the most difficult aspects of second language (L2) acquisition, especially for adult learners (see Flege, Yeni-Komshian, & Liu, 1999; Granena & Long, 2013). This difficulty may be due to the fact that children acquire language-specific patterns of phonemic perception relatively early in the acquisition of their native language (Strange & Shafer, 2008; and see Chládková & Paillereau, 2020 for a review). Once their native phonological system is established, learners do not find it easy to acquire a new phonological system without undue influence from their native language (Casillas, 2020), particularly when the first language (L1) and the L2 possess very different systems (Leather & James, 1991). The current study explored how the acquisition of L2 pronunciation can be facilitated even in the face of vastly different phonological systems. Gestures have frequently been used to enhance pronunciation instruction (see e.g., Acton, 2001; Baker, 2014; Hudson, 2011; Murphy, 2004; Smotrova, 2017), and L2 learners indeed have tended to report that they find these gestures useful (Murphy, 2004; Zheng, Hirata, & Kelly, 2018). However, the effects of instructional gestures on L2 pronunciation have rarely been empirically tested, and the few such reports that do exist did not find positive effects—especially when the instruction targeted individual L2 phonemic features (e.g., Hirata & Kelly, 2010; Hirata, Kelly, Huang, & Manansala, 2014). This lack of empirical support seems surprising considering the positive effects of gesture generally found in other domains (for vocabulary, see Huang, Kim, & Christianson, 2019; Kelly, McDevitt, & Esch, 2009; Macedonia & Klimesch, 2014; Tellier, 2008; for grammar, see Nakatsukasa, 2016, but cf. Nakatsukasa, 2019; for comprehension, see Sueyoshi & Hardison, 2005). Therefore, the present study empirically examined the efficacy of instructional gestures on the development of L2 pronunciation. This study focused on the case of L1 English speakers learning to perceive and produce Japanese. Such learners are assumed to experience difficulties due to the large phonological distance between English and Japanese.

Background Literature

Effects of Instruction on L2 Pronunciation

Studies have reported that most adult L2 learners continue to have pronunciation problems—both in terms of accentedness and comprehensibility—even after they have been immersed in the target language for an extended period of time (e.g., Flege et al., 1999; Saito, 2015). This observation has suggested that simply being exposed to a L2 may not be a sufficient condition for adults to master L2 pronunciation (see DeKeyser & Larson-Hall, 2005). A further implication is that adult L2 learners might benefit from instructional interventions that improve their pronunciation beyond that which they can achieve through exposure alone. Such instructional interventions have been tested in a number of studies, showing overall that L2 learners' phonological knowledge can indeed be improved through instruction. This finding has held in both speech production (e.g., Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Camus, 2019; Saito, 2011, 2013) and perception (e.g., Hirata, 2004; Logan, Lively, & Pisoni, 1991), and in laboratory settings (e.g., Bradlow et al., 1997; Hirata, 2004; Logan et al., 1991; Saito, 2011) and in classroom settings (e.g., Camus, 2019; Saito, 2013; see Saito, 2012 for a review of quasi-experimental studies of L2 pronunciation instruction). The positive effects of instruction have been observed even for notoriously difficult aspects of pronunciation such as the English /r/–/l/ contrast for L1 Japanese learners (e.g., Bradlow et al., 1997; Logan et al., 1991). Lee, Jang, and Plonsky (2015) meta-analysis of 86 studies showed a large effect of instruction on L2 pronunciation, further supporting the overall positive effects of instruction on this domain for adult L2 learners. As further studies were conducted, however, it became evident that the effects of instruction are not without qualification. A more recent meta-analysis showed that the effects of pronunciation instruction appear to be limited to monitored production, such as a read-aloud task, and only to specific phonological features (Saito & Plonsky, 2019). This suggested that L2 learners do not always understand, apply, or remember what is taught through instruction. Therefore, there is a need—for learners and instructors alike—to seek out ways to enhance the efficacy of instruction. One promising technique might be the use of gesture—a tool that has been found to facilitate other domains of L2 instruction (e.g., Allen, 1995; Nakatsukasa, 2016).

Gestures in Teaching Suprasegmentals[1]

Gestures have been widely used in L2 classrooms in an attempt to enhance the effects of pronunciation instruction (see e.g., Hudson, 2011; Smotrova, 2017). For suprasegmental features, language instructors have used, for instance, upward body movements (Smotrova, 2017), finger tapping (Murphy, 2004), handclapping (Baker, 2014), and head nods (Acton, 2001) to visualize and embody word stress and/or the rhythmic patterns of a target language. These gestures highlight L2 features that might otherwise go unnoticed, and teachers (Baker, 2014) and students (Murphy, 2004) have perceived them to be beneficial.

Several experimental studies indeed have appeared to confirm the benefits of gestures for L2 learners' pronunciation of suprasegmental features. Zhang, Baills, and Prieto (2018), for example, found that Chinese monolinguals learned the rhythmic patterns of French words better through the use of handclapping. Metaphoric pitch gestures (Yuan, González-Fuente, Baills, & Prieto, 2019) and rhythmic beat gestures (Gluhareva & Prieto, 2017) have also been reported to help L2 learners improve intonation patterns in Spanish and accentedness in English, respectively. Therefore, when it comes to suprasegmentals, both observational and experimental studies have seemed to suggest that gestures are useful for L2 pronunciation instruction.

Gestures in Teaching Segmentals

Although gestures have been used in L2 classrooms for segmental instruction as well (e.g., Hudson, 2011; Roberge, Kimura, & Kawaguchi, 1996), their benefits do not seem to be as clear as those for suprasegmental instruction (see Kelly, 2017 for a review). One possibility for this dichotomy is that gesture and speech may be connected with higher level features such as suprasegmental information and semantic properties, but they may not be integrated at the lower levels of phoneme processing—making the positive effects of gestures less likely to manifest themselves in segmental instruction (Kelly, 2017). Evidence from experimental studies with Chinese phonemic tone contrasts have seemed to be in line with this suggestion: Virtually no benefit was found when L1 English participants repeated Chinese monosyllables after models with metaphoric gestures depicting the relative height and contour of tones (Zheng et al., 2018). Likewise, Morett and Chang (2015), with similar target stimuli and participants as in the Zheng et al. study, did not find a positive effect of metaphoric gestures on a tone identification test but did find a positive effect on a word-meaning association test, further supporting the idea that gesture may be useful only for learning higher level features such as suprasegmental information and semantic properties (but see also Baills, Suárez-González, González-Fuente, & Prieto, 2019, for a potential positive effect even on tone identification).

In a study related more specifically to ours, the lack of a positive effect of instructional gestures has been reported for Japanese phonemic vowel length contrasts. Hirata and Kelly (2010) used hand movements (short vertical strokes and long horizontal strokes to indicate short and long vowels, respectively) for the Japanese vowel length contrast. In Hirata and Kelly's study, training consisted of L1 English participants' listening to Japanese words containing target vowel contrasts while watching speakers either use or not use hand gestures (among other variables considered in the study). Hirata and Kelly measured learning through an identification task in which participants listened to stimuli and decided whether the second vowel in each stimulus word was short or long. The results did not demonstrate any advantage of gestures over other conditions, suggesting that the gestures did not contribute to the learning of the Japanese vowel length contrast. Building on this study, Hirata et al. (2014) examined the effects of gestures on Japanese vowel length contrasts once more with two modifications. First, in this follow-up study participants not only saw gestures but also performed them, which might have increased the effectiveness of gestures (the enactment effect; Engelkamp & Dehn, 2000). Second, Hirata et al. changed the type of gestures because the original gestures (short vs. long strokes) might not have accurately depicted the structure of the Japanese short versus long contrast. Because a predicted role of gestures in L2 pronunciation instruction is to highlight L2 phonological structure, Hirata et al. (2014) employed moraic-beat gestures that more closely aligned to the underlying structure of the contrast at issue (for details of these two gesture types, see the Instructional Gesture section). Despite these modifications, the second study did not find a positive effect, concluding that "beneficial effects (of gesture) do not seem to exist for segmental phonology" (Hirata et al., 2014, p. 2098).

Although Hirata and colleagues concluded that hand gestures do not appear to contribute to segmental phonology learning, there are still several aspects to consider. First, the participants in their studies were always absolute beginners (with no knowledge of the target language at all). Thus, learners of the language who have some knowledge of its phonological system might be able to make better use of gestures to improve their pronunciation. The findings of research that has involved such learner populations would have more real-world pedagogical implications than studies with complete novices. Second, Hirata and colleagues' studies were conducted in laboratory settings. Because several studies have revealed learners' different behaviors in the classroom and in the laboratory (e.g., Eckerth, 2009; Foster, 1998), Hirata and colleagues' findings should be reexamined in a more ecologically valid classroom setting. Third, Hirata et al.'s (2014) training included a vocabulary learning task as well as a phonological learning task, which might have caused cognitive overload (see e.g., Baills, Zhang, & Prieto, 2018), especially given that the participants were completely new to the language. Thus, there has been a need for studies that focus solely on the phonological property in question to maximize its possible efficacy. Fourth, Hirata and colleagues' studies considered only receptive knowledge of pronunciation. It is, however, important to explore the effects of gestures on productive knowledge as well. Fifth, Hirata and colleagues' studies did not include delayed posttests, and thus the memory-enhancing effect of gestures in the long term (see Allen, 1995; Engelkamp & Dehn, 2000; Nakatsukasa, 2016) remains to be seen in this line of research.

Despite the scarcity of experimental support for the use of gestures in L2 segmental learning (based chiefly on findings by Kelly, Hirata, and their colleagues), gestures have still been used to teach segmental properties along with suprasegmental ones (see e.g., Hudson, 2011). For the Japanese vowel length contrast in particular, accounts from practitioners have reported that producing gestures appears helpful to L2 learners in their production of short and long vowels (Roberge et al., 1996). This has illustrated one instance of the long-standing gap between research and practice regarding gestures and L2 pronunciation instruction. Although gestures have been perceived as being useful in teaching pronunciation (see e.g., Acton, 2001; Baker, 2014; Gilbert, 1978; Hudson, 2011; Murphy, 2004), the literature has lacked robust classroom experimental research to support such effects, especially for segmental features.

Therefore, in this study, we attempted to fill this gap by quantitatively investigating, in a classroom setting, the immediate and delayed effects of gestures on segmental phonological acquisition, focusing on learners' productive knowledge of pronunciation as well as their receptive knowledge. This study explored the following questions:

1. To what extent does seeing and performing gestures facilitate students' development of productive knowledge of L2 Japanese segmental features?
2. To what extent does seeing and performing gestures facilitate students' development of receptive knowledge of L2 Japanese segmental features?

The Current Study

In this study, we examined the effects of instructional gestures on L2 segmental features of pronunciation with a quasi-experimental between-subjects design in which two groups of L1 English learners of beginning Japanese received classroom pronunciation instruction either with or without gestures (gesture group and no gesture group). The linguistic targets consisted of three Japanese segmental features: long vowels, geminates, and moraic nasals. All three of these targets are believed to pose problems for L1 English learners partly due to the discrepancies between their L1 basic phonological unit (the syllable) and the L2 counterpart, which is called the *mora* (Kubozono, 1999). We examined these moraic target features within Japanese loanwords borrowed from English to allow participants to focus on phonological learning by reducing the vocabulary learning aspect of the task (cf. studies that did involve vocabulary learning: Hirata et al., 2014; Kelly & Lee, 2012; Zhang et al., 2018). We selected handclapping, rather than other gesture types, for use in this study for several reasons. First, handclaps are suited to the temporal nature of Japanese—each clap can represent one mora. Further, handclapping has been found in general to be promising in previous studies of L2 pronunciation instruction (Baills et al., 2018; Zhang et al., 2018). Under McNeill's (1992) classification of gestures, handclapping may be considered both a metaphoric gesture and a beat gesture: It visually presents the underlying number of moras in the target word (metaphoric) but also expresses the rhythm of speech in Japanese (beat). Handclapping is also accompanied by sound, making it unique among gesture types. We predicted the current instructional manipulation—students' seeing (and hearing) and performing handclapping—would have a positive impact on the efficacy of instruction by highlighting the target features visually, auditorily, and kinesthetically. This may lead not only to ease of understanding but also to memory enhancement (dual-coding theory: Paivio, 1991; Paivio & Desrochers, 1980; the enactment effect: Engelkamp & Dehn, 2000). To assess these potential effects, we administered productive and receptive knowledge tests to the two groups before, immediately after, and 4 weeks after instruction. We analyzed speech samples from the productive knowledge test based on subjective native speaker ratings and on objective acoustic measures to increase the reliability of the findings. We formulated the following research questions and hypotheses:

- Research Question 1a: To what extent does a gesture group, who sees and produces handclapping during instruction, perform better than a no gesture group, who does not see and produce handclapping during instruction, on a productive knowledge test of Japanese moraic features in terms of native speaker ratings?
- Hypothesis 1a: A gesture group, who sees and produces handclapping during instruction, will perform better than a no gesture group, who does not see and produce handclapping during instruction, on a productive knowledge test of Japanese moraic features in terms of native speaker ratings.
- Research Question 1b: To what extent does a gesture group, who sees and produces handclapping during instruction, perform better than a no gesture group, who does not see and produce handclapping during instruction, on a productive knowledge test of Japanese moraic features in terms of acoustic analysis?

- Hypothesis 1b: A gesture group, who sees and produces handclapping during instruction, will perform better than a no gesture group, who does not see and produce handclapping during instruction, on a productive knowledge test of Japanese moraic features in terms of acoustic analysis.
- Research Question 2: To what extent does a gesture group, who sees and produces handclapping during instruction, perform better than a no gesture group, who does not see and produce handclapping during instruction, on a receptive knowledge test of Japanese moraic features?
- Hypothesis 2: A gesture group, who sees and produces handclapping during instruction, will perform better than a no gesture group, who does not see and produce handclapping during instruction, on a receptive knowledge test of Japanese moraic features.

Method

Participants

We conducted the study at a large university in the southwestern United States. Thirty-one undergraduate students (14 males and 17 females) who were enrolled in either one of two sections of a second-semester beginning Japanese course participated as part of their regular course curriculum. The mean age of the participants was approximately 22 years (range: 18–43; $SD = 4.6$). According to a background questionnaire, all the participants had very limited exposure to the target language outside the classroom, and none of them was a Japanese major.[2] No participants reported any hearing problems or learning disabilities. The main goal of the course was to develop learners' communicative language skills, and the class mostly consisted of speaking practice—participants were therefore accustomed to producing the target language orally in class. The participants had not been specifically taught the pronunciation of the target linguistic features or the concept of mora in class before this study. We arbitrarily assigned each of the two intact classes to one of the two conditions (gesture group, $n = 15$; no gesture group, $n = 16$), but we analyzed the data from only 25 participants (gesture group, $n = 13$; no gesture group, $n = 12$) because three students had a first language other than English (Spanish, $n = 2$; Chinese, $n = 1$),[3] and three other students were absent on the day of instruction.

Target Linguistic Features

The targeted linguistic features, that is, (a) long vowels, (b) geminates, and (b) moraic nasals, uniquely represent the basic phonological unit of Japanese, the mora, and tend to cause learning problems for adult L1 English learners

Table 1 Target Japanese loanwords and their number of moras in the productive knowledge test

Long vowel			Geminate			Moraic nasal		
Target	Meaning	Moras	Target	Meaning	Moras	Target	Meaning	Moras
resur**aa***	wrestler	4	herume**t**to*	helmet	5	to**n**neru*	tunnel	4
juer**ii**	jewelry	4	magune**t**to	magnet	5	cha**n**neru*	channel	4
hiit**aa***	heater	4	pairo**t**to*	pilot	5	fe**n**neru	fennel	4
riid**aa**	leader	4	daie**t**to	diet	5	mado**n**na*	Madonna	4
posut**aa***	poster	4	masuko**t**to*	mascot	5	saba**n**na	Savannah	4
bokus**aa**	boxer	4	tabure**t**to	tablet	5	suza**n**na	Susanna	4

Note. Bold letters indicate target features. Asterisks (*) indicate words that appeared in the instruction session.

of Japanese. It is important to note that Japanese is categorized as a mora-timed language, the native speakers of which produce speech consisting of moras of equal duration (Vance, 1987), which is quite different from English, a stress-timed language (Hoequist, 1983; Kubozono, 1999). This fundamental difference appears to be an obstacle to acquiring L2 systems (see Hirata, 2004, for L1 English speakers' bias toward syllables when counting moras).

Long Vowels

Short and long vowels form minimal pairs in Japanese (e.g., *biru* [building] vs. *biiru* [beer]), although they do not in English. According to Flege (1997), people develop a propensity to classify L2 sounds as equivalent to L1 phonemes after about the age of 5 or 6 years. This established bias toward L1 phonological classification interferes with adult learners' pure sensory perception of L2 sounds (Leather & James, 1991). Therefore, the distinction between short and long vowels in Japanese is difficult for adult L1 English-speaking learners (see Hirata et al., 2014). Thus, we selected long vowels as one of the target linguistic features in the current study. In stress-timed languages like English (Hoequist, 1983), stressed syllables are pronounced with longer duration than unstressed syllables. This property makes it easier for English-speaking learners to perceive and produce long vowels in Japanese loanwords where the long vowel coincides with a stressed syllable in the English equivalent (e.g., *keeki* [cake]). We thus chose as target words loanwords that include unstressed word-final long vowels—which do not provide the benefit of the long vowel corresponding to a stressed vowel in English (see Table 1).

Geminates

A geminate is an obstruent sound that results from doubling a consonant, which produces a mora without a vowel. Just as with long and short vowels, a singleton and a geminate can form a minimal pair in Japanese (e.g., *ite* [stay] vs. *itte* [go]) although they do not in English. For this reason, it is hard for English-speaking learners of Japanese to identify the difference between a singleton and geminate (see Kelly & Lee, 2012), and we thus selected geminates as one of the target linguistic features. Because of the stress-timed nature of English, we chose as target words loanwords in which a geminate does not coincide with a stressed syllable in the English equivalent (see Table 1).

Moraic Nasals

A moraic nasal is a nasal sound that constitutes a mora by itself, without a vowel. Just as with the two other target features, moraic nasals form minimal pairs in Japanese (e.g., *hone* [bone] vs. *honne* [true feelings]) although they do not in English. Although a moraic nasal is not so difficult for nonnative Japanese speakers to identify when it precedes a nonnasal consonant (e.g., *Honda* [a surname]), it is difficult when it is immediately followed by another nasal consonant (e.g., *tannin* [person in charge]) because learners of Japanese can rely only on the duration of the nasal segment to detect its presence (Uchida, 1998). Thus, this feature is also problematic for English-speaking learners of Japanese (see Kariyasu, Ohta, & Snyder, 2007; Tajima, Kato, Rothwell, Akahane-Yamada, & Munhall, 2008), and so we selected moraic nasals as one of the target linguistic features. Just as for long vowels and geminates, we chose as target words loanwords in which a moraic nasal (followed by another nasal consonant) does not coincide with a stressed syllable in the English equivalent (see Table 1).

Instructional Gesture

In the current study, we used handclapping as an instructional gesture. According to Smotrova (2017), effective instructional gestures in teaching pronunciation should (a) visualize and enact pronunciation phenomena, (b) be repetitive, and (c) be synchronously used by both teacher and students. Handclapping appears to be particularly appropriate for the current instructional situation, meeting these criteria: It can visualize and enact Japanese moras by clapping out the equal duration of each mora. Both teacher and students can also use handclapping repetitively and synchronously with relative ease because it is deeply ingrained in many societies; it occurs in situations ranging from early childhood education activities (e.g., Brodsky & Sulkin, 2011) to collective

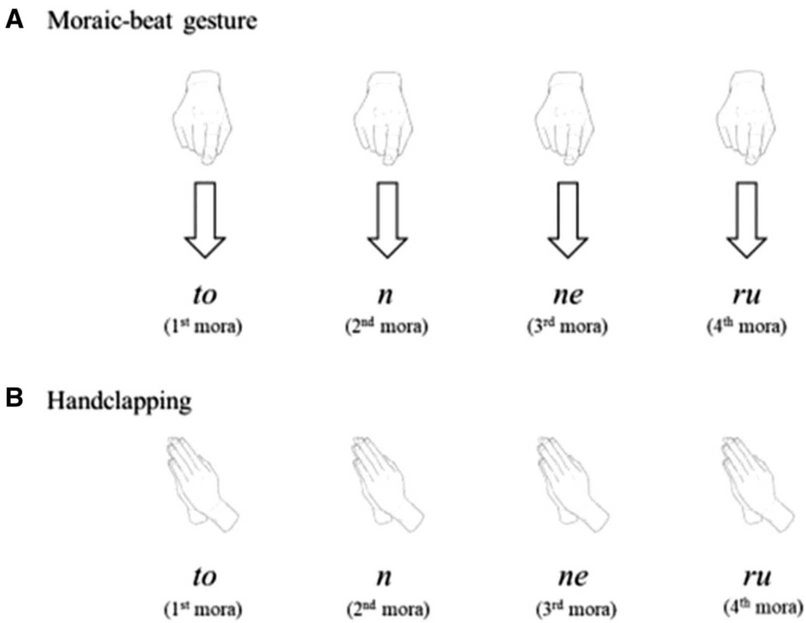

Figure 1 Illustrations of instructional gestures. A four-mora, target word, *tonneru* [tunnel], is shown (the second mora is the target feature, a moraic nasal).

expressions of approval (Néda, Ravasz, Brechet, Vicsek, & Barabási, 2000). In contrast, the type of gestures used in Hirata and Kelly's (2010) study (short vertical strokes and long horizontal strokes to indicate short and long vowels, respectively) might not optimally align with Smotrova's (2017) guidelines: They do not visualize the basic temporal unit of Japanese (the mora) because a single hand stroke could represent either one or two moras, thus not meeting the first criterion. The moraic-beat gesture (chopping downward hand movements; see Panel A in Figure 1) used by Hirata et al. (2014), on the other hand, avoids this problem, yet its practicability in classrooms is unknown because it is not as familiar to students as is handclapping. This perhaps makes handclapping a better choice for synchronous use.[4] For these reasons, we employed handclapping in this study where the instructor and participants in the gesture group clapped their hands to each mora while pronouncing the target words (see Panel B in Figure 1).

Instruction

Instruction consisted of two phases: (a) an explanation of moras and (b) actual pronunciation practice. The instructor first informed the participants that they were going to practice pronouncing Japanese loanwords but did not mention the three target features (i.e., long vowels, geminates, and moraic nasals). In the first phase, the instructor explained to the class the phonological difference between English and Japanese (i.e., Japanese consists of moras, but English consists of syllables) using a PowerPoint slide (about 3 minutes). In the explanation, the instructor showed participants how English words are adopted into Japanese, providing examples. The examples did not include any of the three target linguistic features. Then, the instructor proceeded to the actual pronunciation practice (about 10 minutes), in which participants practiced nine target words (three for each target feature; see Table 1). This second phase consisted of six steps for each target word. Up to Step 4, the instruction was exactly the same in both the gesture and the no gesture groups.

- Step 1: First, the instructor randomly called on a participant, and the participant said a target word in response to an elicitation picture on a PowerPoint slide.
- Step 2: Then, the instructor pronounced the word once.
- Step 3: After that, the instructor randomly called on three participants to guess the number of moras, asking them, "How many moras does it have?"
- Step 4: The instructor then stated the number of moras.
- Step 5: Next, the instructor pronounced the target word three times. In the gesture group, the instructor clapped out each mora while pronouncing the word. In the no gesture group, the instructor stood still without hand (or any other body) movement while pronouncing the word. In both groups, the participants simply watched and listened at this stage.
- Step 6: Then, the instructor once again pronounced the word three times with or without handclapping (as appropriate for each group), but this time the participants repeated the word after the instructor. Participants repeated immediately after each utterance by the instructor (i.e., three turns from the instructor to the students, with each individual turn interweaved with the students' repetition). In the gesture group, participants repeated the instructor's clapping as well.

These steps were repeated for all nine target words. This was a whole class activity, and the elicitation pictures in the practice phase were presented within an overarching story line of traveling to Japan. Each group received the same

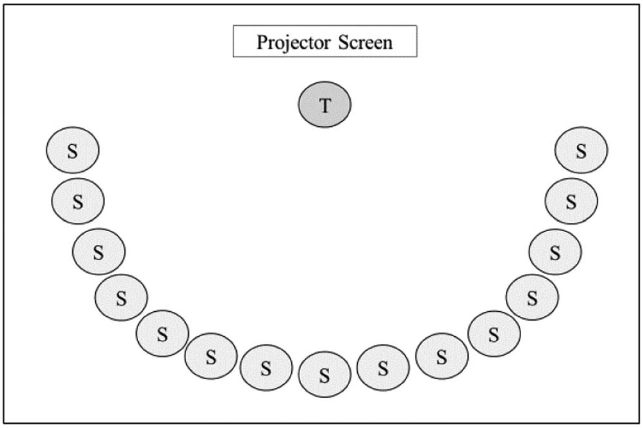

Figure 2 The setting of the classroom. T = teacher. S = student.

Table 2 Instructional design

Phase	Activity
1	Explanation of moras (3 minutes)
2	Pronunciation practice (10 minutes)
	Step 1: First attempt—One participant says the word
	Step 2: Teacher model—Teacher says the word once
	Step 3: Guess—Three participants guess the number of moras
	Step 4: Answer—Teacher says the number of moras
	Step 5: Listen—Participants listen to the word three times (with or without handclapping)
	Step 6: Repeat—Participants listen and repeat the word three times (with or without handclapping)

amount of instruction (about 13 minutes), and the only difference between the two groups was the presence or absence of handclapping in Steps 5 and 6. The PowerPoint presentation used in the instruction had 26 slides in total and was projected on the screen in the classroom (see Figure 2 for the setting of the classroom). The instruction was carried out in English except for the use of the following Japanese phrases: *Kiite kudasai* [Listen to me] and *Minna de itte kudasai* [Repeat after me, everyone], which were familiar to the participants from their regular class sessions. Table 2 summarizes the instructional design.

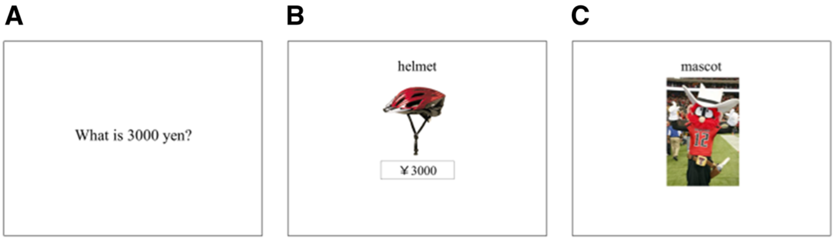

Figure 3 Examples of slides in the productive knowledge test. Panel A shows a question slide in Section 1. Panel B shows a picture slide in Section 1. Panel C shows a picture slide in Section 2. [Color figure can be viewed at wileyonlinelibrary.com]

Assessment

In the current study, we used two assessment instruments: (a) a productive knowledge test in the form of a picture elicitation task and (b) a receptive knowledge test in the form of a dictation task.

Productive Knowledge Test

We used a picture elicitation task to measure participants' production skills. The task elicited the production of 18 target words (six for each target feature; see Table 1). Nine (three for each feature) of the 18 words came from the instructional session, and the remaining nine were new words. We administered this test via a computer in a computer lab. The test had 30 questions in total (18 target items and 12 distractors) and consisted of two sections. We designed all questions so that the participants pronounced loanwords in sentences with the carrier frame, *Kono "loanword" ga …* [This "loanword" is …], so that they would not focus too much on the target words themselves. The first section had 15 questions (nine target items—three for each target feature—and six distractors), in which the participants answered referential questions. They saw a question on the first slide in English (e.g., "What is 3,000 yen?") and then answered the question in Japanese on the basis of the second slide (e.g., *Kono herumetto ga 3,000 en desu*. [This helmet is 3,000 yen.]; see Panels A and B in Figure 3). The second section also had 15 questions (nine target items—three for each target feature—and six distractors), in which the participants answered affective questions. They were instructed to express in Japanese their likes and dislikes about things or people in the pictures (e.g., *Kono masukotto ga suki desu*. [I like this mascot.]; see Panel C in Figure 3). We accompanied all pictures with words in English in order to reduce the possibility of the participants using alternative words. Each section had two practice questions before the

Table 3 Target Japanese loanwords and their number of moras in the receptive knowledge test

Long vowel			Geminate			Moraic nasal		
Target	Meaning	Moras	Target	Meaning	Moras	Target	Meaning	Moras
rait**aa**	lighter	4	kureji**t**to	credit	5	ri**n**neru	Linnell	4
shirub**aa**	silver	4	hariu**d**do	Hollywood	5	ke**n**neru	kennel	4
ood**aa**	order	4	pikuni**k**ku	picnic	5	yoa**n**na	Yoanna	4
bood**aa**	border	4	basuke**t**to	basket	5	iba**n**na	Evanna	4
tesut**aa**	tester	4	boiko**t**to	boycott	5	suza**n**nu	Suzanne	4
misut**aa**	mister	4	berube**t**to	velvet	5	ibo**n**nu	Yvonne	4

Note. Bold letters indicate target features.

actual test questions in order to familiarize participants with the question format. After the directions and practice questions, slides automatically progressed to the next slide at preset intervals (5 seconds for a question slide in Section 1, 10 seconds for a picture slide in Sections 1 and 2, and 5 seconds between questions). The participants spoke into a microphone, and their responses were audio-recorded. We created three versions of the test for the pretest, immediate posttest, and delayed posttest. The elicited loanwords were the same throughout the three versions, but the pictures and elicited sentences were all different, and the order of items was randomized. The Cronbach's alpha estimates of internal consistency for the three versions of this test using native speaker ratings were .77, .76, and .71, respectively, and the Cronbach's alpha estimates for the three versions of the test scored for acoustic duration were .80, .92, and .85, respectively.

Receptive Knowledge Test

We used a dictation task to measure participants' receptive skills. It tested whether they were able to perceive the target linguistic features in 18 words (six for each target feature; see Table 3). None of the words appeared in the instructional session or the productive knowledge test. We administered it as a paper-and-pencil test. The participants listened to the audio and wrote the words in *katakana* or romanization.[5] Each word was uttered in the carrier sentence, *Kore wa "loanword" desu*. [This is "loanword"], with the frame sentence already printed on the test paper, so that participants only had to write the target word. The test had 30 items in total (18 target items and 12 distractors) with two practice items at the beginning. The participants listened to each item twice, once with a male voice and once with a female voice. Two native Japanese speakers had recorded the audio files under the researcher's direction. We asked them to utter each sentence (a) at a natural pace and (b) without any emphasis. We created three versions of the test for the pretest, immediate posttest, and delayed posttest. The items were the same throughout the three versions, but the order of items was randomized. The Cronbach's alpha estimates of internal consistency for the three versions of this test were .81, .92, and .90, respectively.

Procedure

The instructor of one of the two sections of the course, who is the first author of this article, conducted all the instructional sessions and administered the tests. One week before the instruction, both the gesture and no gesture groups took the pretests during their regular class hours in a computer lab. All participants took the productive knowledge test first and then the receptive knowledge test. We administered these tests to all participants simultaneously in each group. One week later, the instructor carried out the instruction for both groups during their regular class hours in the same classrooms where their regular classes were held. We video-recorded the instruction to allow us to check if the participants had followed the instructor's directions.[6] We placed a video camera in the front corner of the classroom to capture both instructor and participants. The instructor informed the participants only that they were going to practice pronouncing Japanese loanwords to enhance their vocabulary and did not mention the specific target linguistic features throughout this experiment. Immediately after the instruction, the participants took the immediate posttests in a computer lab. They also completed the delayed posttests 4 weeks after instruction. Both posttests followed the same procedure as that of the pretests. We administered a background questionnaire in class 1 day after the delayed posttests.

Analysis

Productive Knowledge Test Scoring

In order to make the analysis more valid and credible, we used both subjective ratings by native speakers and objective speech-analysis measures using Praat (Version 6.0; Boersma & Weenink, 2015) to analyze the productive knowledge test data. For this analysis, we carefully clipped each target word utterance from its original sound file, which produced a total of 1,239 sound files.

Subjective measure. We used native speakers' ratings as a subjective measure for the production data. Two native Japanese speakers participated in a rating session. Both were master's students in applied linguistics and were teaching Japanese at a university at the time of this study. The rating session took place individually with the first author in a quiet room. The researcher played each audio file through a speaker. We used a dichotomous scale for this rating. We asked the raters to mark "targetlike" if they thought the participants had pronounced the target linguistic features (long vowels, geminates, and moraic nasals) as native Japanese speakers do, and "non-target-like" if the participants had not pronounced them as native Japanese speakers do. We counted targetlike responses as one point, and non-target-like as zero points. Thus, the participants could receive up to two points for each item (one point contributed by each rater), which made the maximum total score 36 (12 points for each of the three target features). Before they rated each target feature spoken by the participants, we gave the raters eight different examples of a targetlike utterance and eight different examples of a non-target-like utterance. The raters listened to each audio clip of participants' speech only once, but they could listen to a clip again if they were not confident about their judgment of it. In order to avoid bias, we did not inform the raters about which clips were pretest data or posttest data. Each rater scored the audio files in a different order. The rating session lasted about two hours. The interrater reliability measured by Cohen's kappa was .41, which was moderate agreement on the basis of Landis and Koch's (1977) guidelines. Because this value was not particularly high, we recruited two more native speaker raters (lay people residing in Japan) for a follow-up analysis. The interrater reliability of this new set of raters was .62, which was substantial agreement (Landis & Koch, 1977). Because the main findings of the results with these follow-up raters were identical to the ones with the original raters, we have reported the original results (for the results with the follow-up raters, see Appendix S1 in the Supporting Information online).

Objective measure. Using the speech analysis software Praat (Version 6.0; Boersma & Weenink, 2015), we measured the duration of the realization of each target linguistic feature: (a) from the beginning to the end of a target vowel (long vowels), (b) from the end of the preceding vowel to the beginning of the following vowel (geminates), and (c) from the end of the preceding vowel to the beginning of the following vowel (moraic nasals). We calculated each participant's mean duration of each target feature for each test occasion. We divided the mean duration of long vowels by 2, because they consisted of two moras (cf. geminates and moraic nasals, which are one mora each). This produced a score for duration per mora across linguistic features.

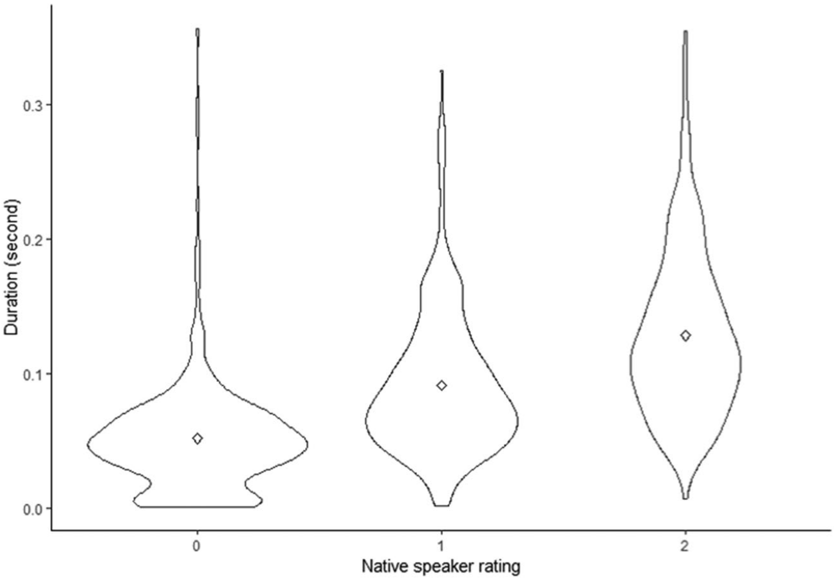

Figure 4 Violin plot of duration of target features by native speaker rating. This data came from both (gesture, no gesture) groups and from all three test points (pretest, immediate posttest, delayed posttest).

We must admit that this duration analysis was not an ideal measure because it was influenced by speech rate.[7] To overcome this limitation, we excluded from the analysis data points that greatly deviated from others (see the Productive Knowledge Test Data Exclusion section).[8] The final duration data correlated largely with native speaker rating scores (polyserial correlation with maximum-likelihood estimate, $\rho = .59$, $SE = .02$, 95% CI [.55, .63]; see also Figure 4), suggesting that the observed longer durations were not just a result of participants' saying the target words slowly.

Receptive Knowledge Test Scoring

The participants received one point if their response demonstrated detection of the target linguistic feature, regardless of orthographical errors or errors in other parts of the word, and they received zero points if their response did not indicate detection of the target feature. The maximum possible score was 18 (six points for each of the three features).

Productive Knowledge Test Data Exclusion

Unfortunately, two participants' data (one for each group) were lost due to technical issues. Also, on the pretest, three participants (one from the gesture group and two from the no gesture group) did not produce the target word in one of the test items. For the calculation of duration for the objective measure, further data elimination was necessary because, in some cases, participants uttered the target words very slowly, which could have skewed the results. Thus, for each item, we excluded values that were 2 standard deviations larger than the mean of the group, which excluded about 4% of the data.[9]

Receptive Knowledge Test Data Exclusion

One participant in the no gesture group scored nearly 80% on the pretest, which stood out from the other scores ($M = 3.79$, or 21%, $SD = 2.40$, 95% CI [2.78, 4.80]). Thus, we excluded this participant from the analysis.

Statistical Procedures

As a preliminary analysis, we carried out independent samples *t* tests to check that the two groups were not significantly different from one another at the time of the pretest. We then performed a mixed design repeated-measures ANOVA on each dependent variable with a between-subjects variable of group (gesture, no gesture) and a within-subjects variable of time (pretest, immediate posttest, delayed posttest). We tested the assumptions of sphericity and normality with Mauchly's test and the Kolmogorov-Smirnov test, respectively. These tests did not indicate any violation of the assumptions (see Appendix S4 for the results of the tests of assumptions). We set the alpha level at .05. We calculated effect size with partial eta squared. Following Cohen's (1988) guideline, we considered an effect size small when η_p^2 was around .01, medium when η_p^2 was around .06, and large when η_p^2 was around .14. When the interaction between time and group was significant, we conducted a post hoc repeated-measures ANOVA for each group to identify different patterns of development. We calculated the effect size for pairwise comparisons with Cohen's *d*. Following the guideline of Plonsky and Oswald (2014) for within-group contrasts, we considered an effect size small when *d* was around 0.60, medium when *d* was around 1.00, and large when *d* was around 1.40. Due to the small number of target items (i.e., six items for each linguistic feature), we conducted these statistical analyses on the total scores of the dependent variables rather than on the separate scores of the linguistic targets.[10] We used SPSS v.22 for the main analyses.

Results

Productive Knowledge Test

Subjective Measure

The first research question concerned the effect of participants' seeing and performing gestures on their productive knowledge of L2 pronunciation. The first analysis tested whether, on the basis of ratings by native speakers of Japanese, the gesture group improved their production of Japanese moraic features over time more than did the no gesture group. Table 4 presents means, standard deviations, and 95% confidence intervals around the means. The two groups were not significantly different from one another at the time of the pretest, $t(21) = -0.62$, $p = .542$, $d = -0.25$, 95% CI (around d) [−1.08, 0.57]. To answer the research question, we performed a mixed design repeated-measures ANOVA for the native speaker rating scores with group (gesture, no gesture) as a between-subjects variable and time (pretest, immediate posttest, delayed posttest) as a within-subjects variable. The results showed that the main effect of time was statistically significant, $F(2, 42) = 31.96$, $p < .001$, $\eta_p^2 = .60$, 90% CI [.42, .69], indicating that, from native raters' point of view, the participants as a whole improved their production of the target features after the instructional session with a large effect size. The pairwise comparisons within the main effect of time revealed that the participants as a whole significantly improved from the pretest to the immediate posttest with a medium effect size, $p < .001$, $d = 1.12$, 95% CI [0.67, 1.65], and that the gain was maintained from the immediate posttest to the delayed posttest (immediate posttest vs. delayed posttest, $p = .541$, $d = 0.10$, 95% CI [−0.24, 0.45]; pretest vs. delayed posttest, $p < .001$, $d = 1.26$, 95% CI [0.75, 1.88]). The interaction effect between time and group was not significant, $F(2, 42) = 2.40$, $p = .103$, $\eta_p^2 = .10$, 90% CI [.00, .23], which suggested that, although the gesture group's scores were numerically higher than those of the no gesture group on the posttests, the use of handclapping during the instructional session did not have a statistically significant impact on the development of participants' productive skills of the target features over time as judged by native speakers of Japanese. The main effect of group was not significant either, $F(1, 21) = 0.21$, $p = .650$, $\eta_p^2 = .01$, 90% CI [.00, .15]. Figure 5 displays these results graphically.

Objective Measure

In the second analysis, we tested whether the gesture group improved their production of Japanese moraic features over time more than did the no gesture group in terms of the duration of the acoustic realization of the target

Table 4 Results for the three assessment measures used in the study

Test/time	All participants ($N = 25$)		Gesture group ($n = 13$)		No gesture group ($n = 12$)	
	M (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI
Productive knowledge test: Subjective native-speaker ratings (maximum score = 36)						
Pretest	15.96 (6.29)	[13.24, 18.68]	15.17 (7.04)	[10.70, 19.64]	16.82 (5.55)	[13.09, 20.55]
Immediate posttest	22.83 (5.47)	[20.47, 25.19]	24.00 (6.28)	[20.01, 27.99]	21.55 (4.34)	[18.63, 24.47]
Delayed posttest	23.39 (4.82)	[21.30, 25.48]	24.33 (5.55)	[20.80, 27.86]	22.36 (3.88)	[19.75, 24.97]
Productive knowledge test: Objective acoustic duration measurements (seconds)						
Pretest	0.22 (0.06)	[0.20, 0.25]	0.22 (0.07)	[0.18, 0.27]	0.22 (0.03)	[0.20, 0.24]
Immediate posttest	0.36 (0.10)	[0.32, 0.41]	0.40 (0.12)	[0.32, 0.48]	0.32 (0.07)	[0.28, 0.36]
Delayed posttest	0.29 (0.08)	[0.25, 0.32]	0.30 (0.08)	[0.25, 0.36]	0.27 (0.07)	[0.22, 0.32]
Receptive knowledge test (maximum score = 18)						
Pretest	3.79 (2.40)	[2.78, 4.80]	3.62 (2.40)	[2.17, 5.06]	4.00 (2.49)	[2.33, 5.67]
Immediate posttest	8.25 (5.37)	[5.98, 10.52]	9.46 (5.56)	[6.10, 12.82]	6.82 (5.00)	[3.46, 10.17]
Delayed posttest	6.00 (4.49)	[4.10, 7.90]	7.31 (4.92)	[4.33, 10.28]	4.46 (3.53)	[2.08, 6.83]

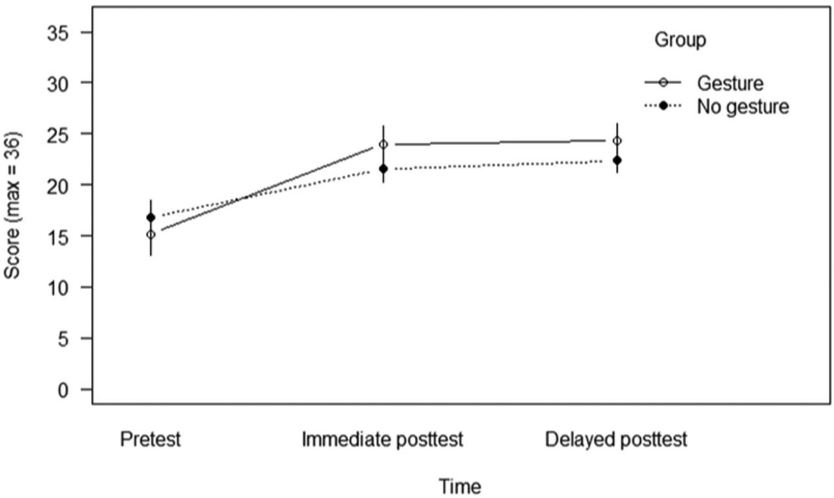

Figure 5 Productive knowledge test results based on native-speaker ratings. The vertical lines represent standard errors.

features. Table 4 presents means, standard deviations, and 95% confidence intervals around the means. The two groups were not significantly different from one another at the time of the pretest, $t(21) = 0.18$, $p = .860$, $d = 0.07$, 95% CI [−0.74, 0.89]. To answer the research question, we performed a mixed design repeated-measures ANOVA for the summed mean target duration with group (gesture, no gesture) as a between-subjects variable and time (pretest, immediate posttest, delayed posttest) as a within-subjects variable. The results showed that the main effect of time was statistically significant, $F(2, 42) = 27.53$, $p <$ .001, $\eta_p^2 = .57$, 90% CI [.37, .66], indicating that the participants as a whole improved their production of the target features after the instructional session with a large effect size. The pairwise comparisons within the main effect of time revealed that the participants as a whole significantly improved from the pretest to the immediate posttest with a large effect size, $p < .001$, $d = 1.58$, 95% CI [0.93, 2.37], and then the gain declined significantly on the delayed posttest, $p < .001$, $d = -0.75$, 95% CI [−1.23, −0.37], but that the duration of the target sounds in their speech was still significantly longer on the delayed posttest than on the pretest, $p = .001$, $d = 0.92$, 95% CI [0.39, 1.51]. The interaction effect between time and group was not significant, $F(2, 42) = 2.12$, $p = .133$, $\eta_p^2 = .09$, 90% CI [.00, .22], which suggested that, although the gesture group's mean duration of the target sounds was numerically longer than

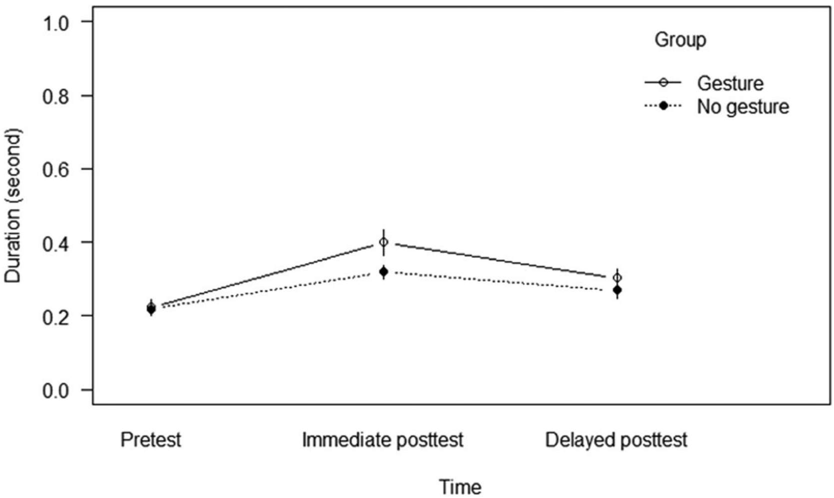

Figure 6 Productive knowledge test results based on objective acoustic duration measurements. The vertical lines represent standard errors.

that of the no gesture group on the posttests, the use of handclapping during the instructional session did not have a statistically significant impact on the development of participants' productive skills of the target features over time. The main effect of group was not significant either, $F(1, 21) = 2.41, p = .136, \eta_p^2 = .10$, 90% CI [.00, .31]. Figure 6 displays these results graphically.

Receptive Knowledge Test

The second research question concerned the effect of participants' seeing and performing gestures on their receptive knowledge of L2 pronunciation. The third analysis tested whether the gesture group improved their perception of Japanese moraic features over time more than did the no gesture group. Table 4 presents means, standard deviations, and 95% confidence intervals around the means. The two groups were not significantly different from one another at the time of the pretest, $t(22) = -0.38, p = .704, d = -0.15$, 95% CI [−0.96, 0.65]. To answer the research question, we performed a mixed design repeated-measures ANOVA for the receptive knowledge test scores with group (gesture, no gesture) as a between-subjects variable and time (pretest, immediate posttest, delayed posttest) as a within-subjects variable. The results showed that the main effect of time was statistically significant, $F(2, 44) = 21.81, p < .001, \eta_p^2 = .50$, 90% CI [.30, .61], indicating that the participants as a whole improved their perception of the target features after the instructional session with a large effect size. The pairwise comparisons within the main effect of time revealed that the participants as a whole significantly improved from the pretest to the immediate posttest with a small-medium effect size, $p < .001$, $d = 0.81$, 95% CI [0.57, 1.56], and then that the gain declined significantly on the delayed posttest, $p < .001$, $d = -0.42$, 95% CI [−0.69, −0.21], but that the scores were still significantly higher on the delayed posttest than on the pretest, $p = .004$, $d = 0.52$, 95% CI [0.18, 1.03]. The interaction effect between time and group was significant, $F(2, 44) = 3.81$, $p = .030$, $\eta_p^2 = .15$, 90% CI [.01, .28], suggesting that the use of handclapping during the instructional session may have had a statistically significant impact over time on the development of participants' receptive skills of the target features with a large effect size. A post hoc repeated-measures ANOVA for each group revealed a significant main effect of time for both the gesture and the no gesture groups, $F(2, 24) = 19.02$, $p < .001$, $\eta_p^2 = .61$, 90% CI [.35, .72] and $F(2, 20) = 6.12$, $p = .008$, $\eta_p^2 = .38$, 90% CI [.07, .54], respectively, suggesting that both groups improved overall. However, the developmental courses were different depending on the use of handclapping. Although both groups improved significantly from the pretest to the immediate posttest (for the gesture group $p < .001$, $d = 0.95$, 95% CI [0.60, 2.10]; for the no gesture group $p = .020$, $d = 0.49$, 95% CI [0.11, 1.29]), and although both groups showed a significant decline on the delayed posttest (for the gesture group $p = .005$, $d = -0.37$, 95% CI [−0.69, −0.12]; for the no gesture group $p = .024$, $d = -0.45$, 95% CI [−1.00, −0.07]), the gesture group still demonstrated a significant gain on the delayed posttest, compared to the pretest, with a small-medium effect size, $p = .004$, $d = 0.74$, 95% CI [0.29, 1.59], whereas the no gesture group was no better than they had been on the pretest at the time of the delayed posttest, $p = .501$, $d = 0.12$, 95% CI [−0.28, 0.57]. The main effect of group was not significant, $F(1, 22) = 1.21$, $p = .283$, $\eta_p^2 = .05$, 90% CI [.00, .24]. Figure 7 displays these results graphically.

Discussion

In the present study, we explored the effects of participants' seeing and performing gestures on their development of receptive and productive knowledge of L2 Japanese segmental features. We had predicted that the use of gestures would facilitate the development of both receptive and productive knowledge. The prediction was confirmed for receptive knowledge, but not for productive knowledge. On the one hand, the results showed that both the gesture and no gesture groups significantly improved their productive knowledge from the

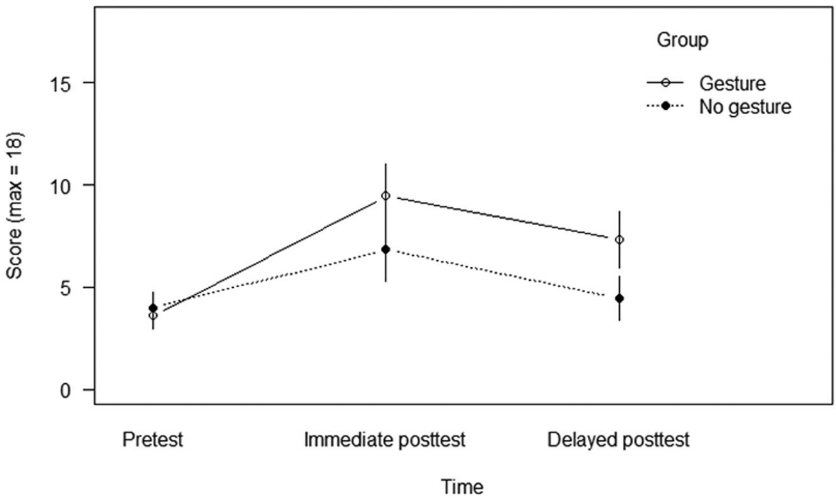

Figure 7 Receptive knowledge test results. The vertical lines represent standard errors.

pretest to the posttests without any significant between-group differences for subjective and objective analyses, suggesting a lack of significant impact of handclapping on productive knowledge. On the other hand, the findings indicate that the two groups went through different patterns of development for receptive knowledge. The gesture group, but not the no gesture group, maintained the instructional effect up to the delayed posttest, suggesting a significant impact of handclapping on receptive knowledge. Why did we find a positive effect of handclapping on receptive knowledge as we expected, but not on productive knowledge? We offer three possible accounts for this discrepancy: (a) skill acquisition theory (DeKeyser, 2015), (b) the learning order of speech (Flege, 1995), and (c) monitoring (e.g., Kormos, 2000).

First, the difference between the receptive and productive knowledge may have resulted from the relationship between the instructional activity and the outcome measures. Skill acquisition theory (DeKeyser, 2015) postulates that knowledge developed through practice is skill-specific, meaning that the acquired skill is useful primarily only in the way that it is practiced (modality, context, etc.; the concept of transfer-appropriate processing could apply here too; Morris, Bransford, & Franks, 1977). Li and DeKeyser (2017), for example, demonstrated that perception-based phonological training led to better performance on a perception test, whereas production-based phonological training led to better performance on a production test, and, critically, benefits were much less pronounced on a test of the reverse (nontrained) skill. This concept can be applied to the present study because the instructional activity—in both groups—mainly focused on production practice with picture stimuli; that is, the instructional activity alone (without handclapping) presumably improved participants' performance on the productive knowledge test for which, similar to the instructional activity, participants had to produce target words in response to pictures. This instructional effect may have been smaller for the receptive knowledge test, which was less similar to the instruction and arguably required more than just the practiced skills alone. The participants, perhaps, needed to extract from the practice the underlying phonological rules to reach the correct answers on the receptive knowledge test. The use of handclapping might have helped this process of noticing and encoding new phonological forms by highlighting the L2 features, thereby possibly making the knowledge more versatile and more observable on receptive as well as productive skills.

However, the learning order of speech may be an alternative explanation for the gap between receptive and productive knowledge. Flege's (1995) speech learning model assumes that the acquisition of perception ability precedes the acquisition of production ability. It is possible that, in this study, handclapping influenced participants' receptive knowledge alone only because it is an earlieracquired domain relative to productive knowledge, and therefore, if a study were to include an extended period of instruction, handclapping might start to influence productive knowledge too.

Yet another possible explanation is the availability of monitoring of L2 form (e.g., Kormos, 2000). As we described in the Methods section, in the productive knowledge test, participants produced target features while answering referential or affective questions, meaning that the participants had to pay attention not only to form but also to meaning. In contrast, on the receptive knowledge test, they were able to focus solely on form. Therefore, it is possible that the advantage afforded by handclapping was only available during testing under participants' conscious monitoring. Thus, we found a positive effect of handclapping in our form-focused, receptive knowledge test but not in our (relatively) meaning-focused, productive knowledge test. This interpretation may gain support from the recent meta-analysis of L2 pronunciation instruction by Saito and Plonsky (2019), which found positive effects of pronunciation instruction only on L2 learners' monitored production, such as on a read-aloud task, with the effects being unclear under more spontaneous production.

On the whole, the present study provides some evidence that gesture in the form of a short intervention of handclapping possibly facilitates the development of L2 segmental features. We assumed that handclapping highlighted the target features visually, auditorily, and kinesthetically, producing this result. Moreover, the fact that we found the effect at the time of the delayed posttest speaks to memory-enhancing effects of multimodal encoding (dual-coding theory; Paivio, 1991; Paivio & Desrochers, 1980, and the enactment effect; Engelkamp & Dehn, 2000). Therefore, this study suggests that the memory-enhancing effect of gestures, which has been found in the domain of vocabulary (e.g., Huang et al., 2019; Kelly et al., 2009; Tellier, 2008), and more recently, in grammar (Nakatsukasa, 2016) and suprasegmental features of speech (e.g., Yuan et al., 2019), might reach the segmental phonology level of L2 acquisition. However, this interpretation must be treated with caution. One reason for this caution is that, as we noted, handclapping provides not only visual and kinesthetic aids, but also provides auditory aids—a characteristic not found in all gesture types. Therefore, the present finding should not be generalized to all types of gesture. For theoretical advancement, further research is needed to find out which aspect(s) of handclapping (visual, kinesthetic, auditory, or a combination of some or all of these) contributes to the positive effect. This inquiry is possible through a laboratory study with experimental manipulations (e.g., muting the sound of handclapping).

Another implication of this study is, perhaps, the importance of including a delayed posttest in this line of research. Neither Nakatsukasa (2016) or the present study found any significant group difference on immediate posttests (i.e., within a day), but both of these studies identified effects on delayed posttests (i.e., over a week). It may be the case that some previous studies (e.g., Hirata & Kelly, 2010; Hirata et al., 2014) did not detect an effect of gestures due to their lack of delayed posttests. In order to better capture the effects of gestures (or any other variable), it would be necessary and desirable to conduct posttests more than once after a reasonable interval.

From a pedagogical point of view, the present study suggests the potential of the use of handclapping to teach L2 pronunciation, indicating small benefits for receptive knowledge. The study also showed no harmful effects for productive knowledge suggesting that clapping did not cause cognitive overload in terms of developing productive skills. By means of visual, auditory, and/or kinesthetic support, handclapping may highlight covert L2 phonological features that might otherwise go unnoticed, possibly making difficult L2 features easier to understand and more memorable. This study demonstrated that this kind of benefit does not appear to be limited to suprasegmentals (e.g., Zhang et al., 2018) but can also be extended to segmentals. The fact that the present study was carried out in a classroom setting, yielding a positive effect of handclapping (at least on receptive knowledge), should also encourage language teachers and learners to actively pursue the use of handclapping and other forms of gesture as an instructional and learning tool, particularly once the current study has generated further empirical support.

Limitations and Future Directions

Although this study has provided some theoretical and pedagogical insights into the use of gestures, it also has some limitations. First, the number of participants may not have been large enough to draw a firm conclusion. Because significance tests are influenced by sample size, the lack of between-group difference on the productive knowledge test in this study (absence of evidence) should not be taken as evidence of absence. Second, the interrater reliability of the native speaker ratings was not particularly high. Although a dichotomous scale was used for the ratings—our reasoning being that native speakers' phonemic perception would be categorical in nature (i.e., native speakers have phonemic boundaries that differentiate one phoneme from another)—this methodological decision appears to have lowered interrater reliability. Therefore, future studies should instead use a more fine-grained scale to allow for more nuanced judgments. Finally, the validity of the dictation task as a measure of perception ability should be discussed. As one reviewer suggested, a multiple-choice recognition task might have been a more accurate measure of perception ability in the sense that participants would not have needed to go through the extra step of writing down what they heard. Despite this drawback, the present study made use of a dictation task over a recognition task to minimize reactivity. Participants might have overemphasized the target segmental features on the productive knowledge test if a recognition pretest had revealed, through contrasting examples, what the researchers were investigating. This in turn might have unduly hidden L1 English speakers' typical production errors for these particular features (i.e., short vowel instead of long vowel; singleton instead of geminate; lack of moraic nasal). This was also part of the reason that the target features were not mentioned to participants during the study. All this being said, the accuracy of outcome measures is an important aspect to consider, and future researchers need to choose tests on the basis of their research purposes and designs.

Conclusion

The present study aimed to advance research on the pedagogical use of gestures for pronunciation instruction by quantitatively exploring the immediate and delayed effects of handclapping in classrooms. The results indicate that gestures may indeed be useful by making difficult L2 phonological features easier to understand and more memorable. This study is only a first step, and many more studies are needed, especially considering that L2 teachers frequently use gestures to teach pronunciation (see e.g., Hudson, 2011; Smotrova, 2017) without substantial empirical support for the practice. Therefore, future research should continue to explore the impact of gestures in classrooms and accumulate findings so that language teachers and learners can systematically use gestures as powerful instructional and learning tools.

Final revised version accepted 8 February 2020

Notes

1 Following Hirata, Kelly, Huang, and Manansala (2014) and Zheng, Hirata, and Kelly (2018), we used the terms *segmental* and *suprasegmental*, with segmental referring to acoustic features within a segment (e.g., individual phonemic contrasts) and suprasegmental referring to acoustic features beyond a segment (e.g., sentential intonation).

2 The current sample of students included various majors with psychology accounting for the largest proportion ($n = 6$), but none of the participants' majors seemed to be of potential relevance to this study except one French major. A follow-up analysis with this participant removed did not change any of the main findings.

3 A reviewer commented that we should also have considered knowledge of other L2s for inclusion criteria. We did not collect this information, so we were not able to pursue this variable in the present study. Future studies should include such information for more complete analysis.

4 In fact, our pilot study with the moraic-beat gesture resulted in many participants unable to properly repeat this gesture after the instructor.

5 In Japanese, loanwords are written in a distinct writing system, called *katakana*. However, because of the limited proficiency of the participants in the current study, we allowed participants to use romanization instead of *katakana*. Otherwise, orthographical issues might have obscured the results.

6 Occasional video-recording of the classroom was mentioned in the course syllabus, and this was one of the several occasions when the class was video-recorded.

7 To minimize the influence of speech rate on objective analysis, other analysis methods and research designs are also possible. For instance, if the research materials include minimal pairs (e.g., short vs. long vowels), direct pairwise comparisons are possible. Because the Japanese language consists of moras, each of which has equal duration (Vance, 1987), a reasonable alternative might also be to use ratio calculation (i.e., the ratio of the duration of a target feature to that of the entire word).

8 A reviewer suggested converting the data to *z* scores for analysis and trimming them as necessary. We conducted follow-up analyses with this method. The main result of interest was the same as in the original analysis: There was no significant difference between the two groups for this dependent variable (see Appendix S2 in the Supporting Information online for the details).

9 A reviewer expressed concern about excluding 4% of the data. We conducted a follow-up analysis with all data points included. The main result of interest was the same as in the original analysis: There was no significant difference between the two groups for this dependent variable (see Appendix S3 in the Supporting Information online for the details).

10 Although we thought that the number of items was too small for separate analyses of different linguistic targets or of trained (i.e., used in the instruction) versus untrained items, we conducted these analyses upon the reviewers' request. We performed repeated-measures MANOVAs with dependent variables of different linguistic targets (long vowels, geminates, moraic nasals) and different training status (trained items, untrained items). The results did not yield new significant findings for these variables.

Open Research Badges

This article has earned an Open Materials badge for making publicly available the components of the research methods needed to reproduce the reported procedure. All materials that the authors have used and have the right to share are available at https://www.iris-database.org. All proprietary materials have been precisely identified in the manuscript.

References

Acton, W. (2001). FocalSpeak: Integrating rhythm and stress in speech-pronunciation. In J. Murphy & P. Byrd (Eds.), *Understanding the courses we teach: Local perspectives on English language teaching* (pp. 197–217). Ann Arbor: University of Michigan Press. https://doi.org/10.3998/mpub.11352

Allen, L. Q. (1995). The effects of emblematic gestures on the development and access of mental representations of French expressions. *The Modern Language Journal*, *79*, 521–529. https://doi.org/10.1111/j.1540-4781.1995.tb05454.x

Baills, F., Suárez-González, N., González-Fuente, S., & Prieto, P. (2019). Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Studies in Second Language Acquisition*, *41*, 33–58. https://doi.org/10.1017/S0272263118000074

Baills, F., Zhang, Y., & Prieto, P. (2018). Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: Evidence from Catalan and Chinese learners of French. In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), *Proceedings of the 9th international conference on speech prosody* (pp. 853–857). Poznań, Poland: International Speech Communication Association. https://doi.org/10.21437/SpeechProsody.2018-172

Baker, A. (2014). Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *TESOL Quarterly*, *48*, 136–163. https://doi.org/10.1002/tesq.99

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer [Computer software]. Retrieved from http://praat.org

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*, 2299–2310. https://doi.org/10.1121/1.418276

Brodsky, W., & Sulkin, I. (2011). Handclapping songs: A spontaneous platform for child development among 5–10-year-old children. *Early Child Development and Care*, *181*, 1111–1136. https://doi.org/10.1080/03004430.2010.517837

Camus, P. (2019). The effects of explicit pronunciation instruction on the production of second language Spanish voiceless stops. *Instructed Second Language Acquisition*, *3*, 81–103. https://doi.org/10.1558/isla.37279

Casillas, J. (2020). The longitudinal development of fine-phonetic detail: Stop production in a domestic immersion program. *Language Learning*, *70*, https://doi.org/10.1111/lang.12392

Chládková, K., & Paillereau, N. (2020). The what and when of universal perception: A review of early speech sound acquisition. *Language Learning*, *70*

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.) Hillsdale, NJ: Lawrence Erlbaum.

DeKeyser, R. M. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 94–112). New York, NY: Routledge.

DeKeyser, R., & Larson-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 88–108). New York, NY: Oxford University Press.

Eckerth, J. (2009). Negotiated interaction in the L2 classroom. *Language Teaching*, *42*, 109–130. https://doi.org/10.1017/S0261444808005442

Engelkamp, J., & Dehn, D. M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 671–682. Retrieved from https://psycnet.apa.org/doi/10.1037/0278-7393.26.3.671

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E. (1997). English vowel productions by Dutch talkers: More evidence for the "similar" vs "new" distinction. In A. James & J. Leather (Eds.), *Second-language speech: Structure and process* (pp. 11–52). Berlin, Germany: De Gruyter Mouton.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*, 78–104. https://doi.org/10.1006/jmla.1999.2638

Foster, P. (1998). A classroom perspective on the negotiation of meaning. *Applied Linguistics*, *19*, 1–23. https://doi.org/10.1093/applin/19.1.1

Gilbert, J. (1978). Gadgets: Some non-verbal tools for teaching pronunciation. *TESL Reporter*, *11*(2), 6–13.

Gluhareva, D., & Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research*, *21*, 609–631. https://doi.org/10.1177/F1362168816651463

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*, 311–343. https://doi.org/10.1177/F0267658312461497

Hirata, Y. (2004). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *The Journal of the Acoustical Society of America*, *116*, 2384–2394. https://doi.org/10.1121/1.1783351

Hirata, Y., & Kelly, S. D. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, *53*, 298–310. https://doi.org/10.1044/1092-4388(2009/08-0243)

Hirata, Y., Kelly, S. D., Huang, J., & Manansala, M. (2014). Effects of hand gestures on auditory learning of second-language vowel length contrasts. *Journal of Speech, Language, and Hearing Research*, *57*, 2090–2101. https://doi.org/10.1044/2014_JSLHR-S-14-0049

Hoequist, C. Jr. (1983). Syllable duration in stress-, syllable- and mora-timed languages. *Phonetica*, *40*, 203–237. https://doi.org/10.1159/000261692

Huang, X., Kim, N., & Christianson, K. (2019). Gesture and vocabulary learning in a second language. *Language Learning*, *69*, 177–197. https://doi.org/10.1111/lang12326

Hudson, N. (2011). *Teacher gesture in a post-secondary English as a second language classroom: A sociocultural approach* (Doctoral dissertation). Retrieved from https://digitalscholarship.unlv.edu/thesesdissertations/1044/

Kariyasu, M., Ohta, E., & Snyder, S. (2007). Nihongo to eigo chuugokugo tono onseigakutekitokuchou to souiten [Phonetic features and notable differences between Japanese, English and Chinese]. *Journal of Kyushu University of Health and Welfare*, *8*, 133–138.

Kelly, S. D. (2017). Exploring the boundaries of gesture–speech integration during language comprehension. In R. B. Church, M. W. Alibali, & S. D. Kelly (Eds.), *Why gesture? How the hands function in speaking, thinking and communicating* (pp. 243–265). Amsterdam, Netherlands: John Benjamins. https://doi.org/10.1075/gs.7

Kelly, S. D., & Lee, A. L. (2012). When actions speak too much louder than words: Hand gestures disrupt word learning when phonetic demands are high. *Language and Cognitive Processes*, *27*, 793–807. https://doi.org/10.1080/01690965.2011.581125

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*, 313–334. https://doi.org/10.1080/01690960802365567

Kormos, J. (2000). The role of attention in monitoring second language speech production. *Language Learning*, *50*, 343–384. https://doi.org/10.1111/0023-8333.00120

Kubozono, H. (1999). Mora and syllable. In N. Tsujimura (Ed.), *The handbook of Japanese linguistics* (pp. 31–61). Oxford, UK: Blackwell. https://doi.org/10.1002/9781405166225.ch2

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. Retrieved from https://www.jstor.org/stable/2529310

Leather, J., & James, A. (1991). The acquisition of second language speech. *Studies in Second Language Acquisition*, *13*, 305–341. https://doi.org/10.1017/S0272263100010019

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*, 345–366. https://doi.org/10.1093/applin/amu040

Li, M., & DeKeyser, R. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Acquisition*, *39*, 593–620. https://doi.org/10.1017/S0272263116000358

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*, 874–886. https://doi.org/10.1121/1.1894649

Macedonia, M., & Klimesch, W. (2014). Long-term effects of gestures on memory for foreign language words trained in the classroom. *Mind, Brain, and Education*, *8*(2), 74–88. https://doi.org/10.1111/mbe.12047

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

Morett, L. M., & Chang, L. Y. (2015). Emphasising sound and meaning: Pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience*, *30*, 347–353. https://doi.org/10.1080/23273798.2014.923105

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Murphy, J. (2004). Attending to word-stress while learning new vocabulary. *English for Specific Purposes*, *23*, 67–83. https://doi.org/10.1016/S0889-4906(03)00019-X

Nakatsukasa, K. (2016). Efficacy of recasts and gestures on the acquisition of locative prepositions. *Studies in Second Language Acquisition*, *38*, 771–799. https://doi.org/10.1017/S0272263115000467

Nakatsukasa, K. (2019). Gesture-enhanced recasts have limited effects: A case of the regular past tense. *Language Teaching Research*, 136216881987028, https://doi.org/10.1177/1362168819870283

Néda, Z., Ravasz, E., Brechet, Y., Vicsek, T., & Barabási, A. L. (2000). Self-organizing processes: The sound of many hands clapping. *Nature*, *403*, 849–850. https://doi.org/10.1038/35002660

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*, 255–287. Retrieved from https://psycnet.apa.org/doi/10.1037/h0084295

Paivio, A., & Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology*, *34*, 388–399. Retrieved from https://psycnet.apa.org/doi/10.1037/h0081101

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912. https://doi.org/10.1111/lang.12079

Roberge, C., Kimura, M., & Kawaguchi, Y. (1996). *Nihongo no Hatsuon Shidoo: VT-hoo no Riron to Jissai* [Pronunciation training for Japanese: Theory and practice of the VT method]. Tokyo, Japan: Bonjinsha.

Saito, K. (2011). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: An instructed SLA approach to L2 phonology. *Language Awareness*, *20*, 45–59. https://doi.org/10.1080/09658416.2010.540326

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, *46*, 842–854. https://doi.org/10.1002/tesq.67

Saito, K. (2013). Reexamining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition*, *35*, 1–29. https://doi.org/10.1017/S0272263112000666

Saito, K. (2015). The role of age of acquisition in late second language oral proficiency attainment. *Studies in Second Language Acquisition*, *37*, 713–743. https://doi.org/10.1017/S0272263115000248

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*, 652–708. https://doi.org/10.1111/lang.12345

Smotrova, T. (2017). Making pronunciation visible: Gesture in teaching pronunciation. *TESOL Quarterly*, *51*, 59–89. https://doi.org/10.1002/tesq.276

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini

(Eds.), *Phonology and second language acquisition* (pp. 153–192). Amsterdam, Netherlands: John Benjamins. https://doi.org/10.1075/sibil.36

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, *55*, 661–699. https://doi.org/10.1111/j.0023-8333.2005.00320.x

Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., & Munhall, K. G. (2008). Training English listeners to perceive phonemic length contrasts in Japanese. *The Journal of the Acoustical Society of America*, *123*, 397–413. https://doi.org/10.1121/1.2804942

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, *8*, 219–235. https://doi.org/10.1075/gest.8.2.06tel

Uchida, T. (1998). Nihongo tokushuhaku no shinritekina ninchikatei kara toraeta onsetsu to haku [Categorical perception of relatively steady-static speech sound duration in Japanese moraic phonemes]. *Journal of the Phonetic Society of Japan*, *2*(3), 71–86.

Vance, T. J. (1987). *An Introduction to Japanese phonology*. Albany: State University of New York Press.

Yuan, C., González-Fuente, S., Baills, F., & Prieto, P. (2019). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*, *41*, 5–32. https://doi.org/10.1017/S0272263117000316

Zhang, Y., Baills, F., & Prieto, P. (2018). Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: Evidence from training Chinese adolescents with French words. *Language Teaching Research*. https://doi.org/10.1177/F1362168818806531

Zheng, A., Hirata, Y., & Kelly, S. D. (2018). Exploring the effects of imitating hand gestures and head nods on L1 and L2 Mandarin tone production. *Journal of Speech, Language, and Hearing Research*, *61*, 2179–2195. https://doi.org/10.1044/2018_JSLHR-S-17-0481

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Follow-Up Native Speaker Rating Analysis with Another Set of Raters.
Appendix S2. Follow-Up Duration Analyses with *z* Scores.
Appendix S3. Follow-Up Duration Analysis with All Data Points.
Appendix S4. Results of Tests of Assumptions.

Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

Gesture May Have a Memory Enhancing Effect in the Teaching of Sounds in a Second Language

What This Research Was About and Why It Is Important

Language teachers use a wide range of instructional gestures, including body movements, head nods, and handclapping in second language (L2) classrooms to help learners to master L2 pronunciation. However, the effects of such techniques have rarely been tested. This study examined one form of instructional gesture, handclapping, in L2 Japanese classrooms to see whether and how it facilitated the development of L2 sounds. Two groups of English-speaking learners of Japanese practiced pronouncing sounds that are unique to the L2 either with or without handclapping. The results showed that those who practiced with handclapping better maintained the effects of instruction in a perception test than those without handclapping, suggesting a memory enhancing effect of the technique.

What the Researchers Did

- Twenty-five English-speaking university students in a second-semester Japanese course received 10 minutes of pronunciation instruction in which they practiced pronouncing unique Japanese sounds either with (13 students) or without (12 students) handclapping.
- The learners also took production and perception tests at three time points: before instruction, immediately after instruction, and 4 weeks after instruction.
- The production test data were analyzed in both subjective and objective ways: scoring by native Japanese speakers and by acoustic analysis using a software.

What the Researchers Found

- On the production tests:
    - As measured by native Japanese speakers' ratings, learners who practiced with handclapping performed similarly to their counterparts without handclapping.
    - As measured by the objective acoustic analysis, learners who practiced with handclapping performed similarly to their counterparts without handclapping.

  - These findings suggest there may be no harmful effects of handclapping for production.
- On the perception tests:
  - Handclapping helped the learners to maintain a small–medium sized benefit for receptive knowledge 4 weeks after instruction, whereas practice without handclapping did not.

Things to Consider

- The study found small–medium benefits of handclapping for perception and no disadvantages for production.
- Instructional gestures may have a memory enhancing effect on the teaching of L2 sounds, and therefore language teachers could consider incorporating such techniques into instruction.
- However, future studies are needed to see whether other forms of gestures (aside from handclapping) have a similarly positive effect and whether the benefit could be extended to learners' production skills.
- This study had small numbers of participants and so results should be interpreted with caution.

Materials and data: Materials are publicly available at: https://www.iris-database.org

How to cite this summary: Iizuka, T., Nakatsukasa, K., & Braver, A. (2020). Gesture may have a memory enhancing effect in the teaching of sounds in a second language. *OASIS Summary* of Iizuka, Nakatsukasa, & Braver (2020) in *Language Learning*. https://oasis-database.org