To be published in *System (*Elsevier)

The Effects of Perception- vs. Production-based Pronunciation Instruction

Bradford Lee

Luke Plonsky

Kazuya Saito

Abstract

While research has shown that provision of explicit pronunciation instruction (PI) is facilitative of various aspects of second language (L2) speech learning (Thomson & Derwing, 2015), a growing number of scholars have begun to examine which type of instruction can best impact on acquisition. In the current study, we explored the effects of perception- vs. production-based methods of PI among tertiary-level Japanese students of English. Participants (N = 115) received two weeks of instruction on either segmental or suprasegmental features of English, using either a perception- or a production-based method, with progress assessed in a pre/post/delayed posttest study design. Although all four treatment groups demonstrated major gains in pronunciation accuracy, performance varied considerably across groups and over time. A close examination of our findings suggested that perception-based training may be the more effective training method across both segmental and suprasegmental features.

Keywords: L2 pronunciation, pronunciation instruction, perception-based instruction, production-based instruction, English as a foreign language

1. Introduction

The processes involved in learning a second language (L2) include much more than memorizing a new set of vocabulary and a few syntactic and grammatical rules. A number of scholars have tested, confirmed, and emphasized the importance of pronunciation, stating that effective communication is impossible when nonnative speakers' pronunciation falls below a certain threshold level, even when their vocabulary and grammar are excellent (Derwing & Munro, 2015; Levis, 2018). To date, there is ample empirical evidence that provision of explicit phonetic instruction is facilitative of various dimensions of L2 pronunciation development (Lee, Jang, & Plonsky, 2015; Saito & Plonsky, 2019; Thomson & Derwing, 2015). What has remained open to debate, however, is what type of instruction can help L2 learners acquire new sounds in the most effective and efficient manner. In the context of 115 Japanese learners of English, the current study is designed to examine the relative impact of two different types of instructional approaches—i.e., whether the instruction is perception-based (i.e., aimed at increasing the participants' identification or discrimination abilities) or production-based (i.e., eliciting the correct articulation of the target features while making use of corrective feedback). Examining this topic is directly relevant not only to designing an optimal pedagogy for a particular group of L2 learners (inexperienced English-as-a-Foreign-Language [EFL] learners in Japan), but also to testing existing theoretical controversies on the nature, directionality, and transferability of the link between the perception and production dimensions of L2 instruction (Shintani, Li, & Ellis, 2013).

1.1. Literature Review

Theoretical frameworks for pronunciation instruction (PI) can be categorized as either being perception- or production-based. This broad categorization is based on whether comprehension or articulation of the target feature is viewed as the source of acquisition (Shintani et al., 2013). This debate has long existed in Instructed Second Language Acquisition (which considers the effects of active interventions), as production-based instruction methodologies stretch as far back as the Grammar Translation and Audiolingual Methods which remain in practice to the present day (Heinz, 2013; Mart, 2013; Ellis, 2003). Some studies have offered empirical evidence that production-based methodologies such as explicit feedback (Ellis, 2001) and prompts (Lyster, Saito, & Sato, 2013) which elicit production from students in communicatively authentic contexts, have the potential to increase students' mental representations and processing ability of the target feature. Meanwhile, proponents of comprehension-based instruction methods (i.e., perception) have also produced empirical studies which suggest that it is in fact superior. Ellis' (1997) computation model suggests that L2 acquisition begins when learners are able to consciously notice linguistic information they are exposed to in L2 input. From this point of view, if learners are able to convert *input* into *intake*, they will proceed to internalize the information, restructure their interlanguage, and create form-meaning mapping. According to this model, teachers can assist their students' interlanguage development by drawing their attention to problematic features and increasing their frequency or

salience to encourage noticing (for more on noticing, see also Schmidt, 1990, Lyster, 2007; VanPatten, 2007).

Specifically relating to pronunciation, several theoretical models have been proposed which reflect the ongoing debate as to whether perception or production is the source of acquisition. Production-based models such as the Markedness Differential Hypothesis (Eckman, 1977, 1991) or the Ontogeny Model (later the Ontogeny Phylogeny Model) (Major, 1986, 2001) posit that a learner's pronunciation success or failure is based largely on the difference in markedness between the L2 and the speaker's native language (L1) and are independent of comprehension ability. In contrast, Best's Perceptual Assimilation Model (PAM) (Best, 1994, 1995) theorizes that a person's perceptual system gradually becomes exclusively attuned to L1-specific sounds and thus progressively worse at discerning sounds that are not contrastive in the L1. Best claims that the mechanism by which this happens is based on phonetic-articulatory patterning, where a listener subconsciously filters nonnative speech sounds based on the manner and place of articulation and maps (categorizes) where they fall along a gradient. The L1 influences the ability of a learner to discern the level of contrastive phonetic detail in nonnative speech input (cf. Best & Tyler, 2007).

Perhaps the most well-known perception-based model is Flege's Speech Learning Model (SLM) (Flege, 1988, 1992, 1995, 1999, 2002). The SLM was designed specifically with the L2 learner in mind, attempting to describe the process by which a learner perceives then comes to produce L2 speech sounds. This theoretical model is directly relevant to the current study. If we follow the main tenet of the SLM (i.e., L2 speech learning is perception-based), it is reasonable to assume that guiding students to improve their perception rather than production ability may maximize the process and product of acquisition under classroom conditions.

The SLM is perhaps the strongest of the perception-based theories, claiming that perception is a necessary precursor to production. In order to be able to produce an L2-specific sound, a speaker must first be able to detect novel sounds in the L2 and create a corresponding mental representation of it (i.e., a phonetic category). The ability of a learner to do this limits the degree to which L2 sounds may be produced. Flege (2007) states:

> …an L2 learner will at first use the closest L1 sound to produce L2 sounds, without evidence of modification of learning. … The SLM predicts that, when L2 category formation is blocked, production of an L2 sound will be modified slowly over time if the L2 sound differs audibly from the L1 sound with which it has been equated. (p. 368)

Unlike other perception-based models, the SLM postulates that "the mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning" (Flege, 1995, p. 239).

In the context of Japanese learners of English (the target population of this study), a number of studies have clearly shown that intensive perception training without any explicit production instruction can result in improvement in production (Saito, 2015; Bradlow, Pisoni,

Akahane-Yamada, & Tohkura, 1997; see also Saito & Wu, 2014 for similar results on suprasegmental targets). There is evidence to suggest that L2 learners tend to have better ability in perception than production (Baker & Trofimovich, 2006; Flege, 1993; Flege, Bohn, & Jang, 1997). While the SLM predicts that production ability cannot exceed perception ability, some cases where perception has increased with no corresponding detectable gains in production have also been reported (Aliaga-Garcia & Mora, 2009; Munro, 1993).

Sakai and Moorman's (2018) meta-analysis specifically sought to address this discrepancy in results. Overall, their results showed a small- to medium-sized relationship between perception and production gains. That is, perception-only training led not only to improvements in perception (d = .92) but to production as well (d = .52). The authors attribute the difference in effects on auditory-to-articulatory mapping to several variables, such as target phoneme, learning context, amount of experience in the L2, individual review/training, training duration, and the presence or absence of phonetic instruction.

1.2. Research Question

Flege's SLM posits that learners must be able to audibly perceive the differences between L2 speech sounds and their L1 before increases in production accuracy can occur. This implies that learners' attention must be explicitly drawn to the differences in the L2 and the L1 via form-focused instruction (FFI), and that errors in the learners' L2 production would benefit from explicit corrective feedback. This study will test and compare the effects of such explicit PI using either perception- or production-based FFI. As per Saito and Plonsky's (2019) synthesis of PI literature, perception-based FFI is designed to help learners grasp the perceptual similarities and dissimilarities between L2 sounds and L1 counterparts. In production-based FFI, teachers guide students to understand the manner and place of articulation of L2 consonants and vowels relative to that of L1 counterparts and how to use such articulators.

To date, while the relevant speech learning theories were mainly intended to account for differences in the learnability of individual phonemes, it is important to note here that some attempts have been made to extend those accounts to suprasegmental features (e.g., So & Best, 2010 for PAM-L2; Trofimovich & Baker, 2006 for SLM). Building on this line of work, the current investigation also seeks to examine whether the model could be generalized to suprasegmental features as well. This study will therefore seek to answer the following research question (RQ):

What are the relative effects of perception-based versus production-based training of syllables and phonemes on pronunciation accuracy?

It should be noted that the focus of this study was whether the SLM can accurately predict which instruction method would result in the greatest acquisition of the target segmental and suprasegmental features. Pronunciation accuracy was therefore defined as the successful production of the Standard American English (SAE) features taught in the treatments. This is not meant to imply that non-SAE pronunciation is in any way erroneous, simply the production of such features reflects the inability of the students to acquire the target sounds. As the SLM posits

that L2 pronunciation acquisition takes place when learners are able to hear new sounds (Flege, 1995), it was predicted that perception-based training would be more effective than production-based training. However, what makes this study unique is that we attempted to test the SLM's assumptions beyond the traditional segmental context and onto suprasegmental training as well. Accordingly, we hypothesized that the advantage favoring perception-based training would be observable in both segmental and suprasegmental training contexts. Previous studies have found similarities in L2 speech learning patterns at segmental and suprasegmental levels (e.g., Trofimovich & Baker, 2006); this study sought to further extend these findings to the area of pronunciation accuracy.

2. Method

2.1. Participants

This study was conducted in an EFL setting at a small university in rural Japan, where participants had little chance to use English in communicative contexts outside of the classroom. As most of their studies had been geared towards developing receptive skills useful in test-taking such as listening-for-gist and reading comprehension, the students' oral proficiency level could be described as low-intermediate.

Recruitment efforts resulted in a total of 119 participants. However, four members either chose to drop out or were unable to attend one or more of the training sessions. These individuals' data were excluded from the final analyses, resulting in a final sample of $N = 115$. The participants were randomly divided into five groups: control group (CG; $n = 23$), syllabic perception instruction (SPe; $n = 21$), syllabic production instruction (SPr; $n = 22$), phonemic perception instruction (PPe; $n = 24$), and phonemic production instruction (PPr; $n = 25$). The control group received no treatment for the duration of the study, apart from their normal coursework at the university, and were only asked to convene for the pre- and posttests. Although the duration of the study was relatively short (a total of 4 weeks from pretest to delayed posttest), and normal classwork at the university does not consist of any PI, the establishment of the control group was necessary to determine if any gains demonstrated by the experimental groups could be attributed solely to the treatments, or if there were other factors (e.g., test practice effects) that needed to be considered.

2.2. Instruments

2.2.1. Pronunciation accuracy assessments.

In order to capture the multifaceted nature of instructed L2 pronunciation development (consolidating, proceduralizing, and automatizing what participants have learned from PI), controlled and spontaneous production tasks were devised as outcome measures. The former task is designed to elicit more explicit, analyzed, and conscious knowledge of L2 pronunciation forms; and the latter is designed to measure L2 learners' relatively unconscious and unmonitored use of L2 pronunciation forms (for methodological discussion on task modality in PI research, see Saito & Plonsky, 2019). A PowerPoint slideshow of 30 slides was designed to test 10 English words three times each: in a free-response style question, a direct translation task from Japanese

to English, and finally a read-aloud word list. To ensure comprehension and reduce anxiety, all question prompts and instructions were written in the participants' L1 (Japanese).

Prior to the start of the experiment, a total of three piloting sessions were conducted by the lead researcher with the help of eight bilingual L1 Japanese individuals of similar age and ability as the target population. The goal of piloting was to a) identify problematic pronunciation areas for the sample population, b) confirm participants could understand the instructions, and c) confirm that the language being elicited contained the desired target words. For illustration, during the piloting phase "orange juice" was included as a target word, so a picture of an orange-colored liquid in a glass was shown to participants with the prompt, この飲み物は何ですか？(*What drink is this?*). However, some participants responded, "This drink is juice", resulting in a change of the visual aid and the target word to "orange" in the final version[1]. The target words and the features they were designed to assess are noted below in Table 1.

Table 1

Target Words and Observed Problematic Features at the Piloting Phase

Word	Standard vs. Observed Pronunciation	Problematic Issue(s)
milk	[mɪlk] → [ˈmɪluku]	anaptyxis
bath	[bæθ] → [bæs]	phonemic
chocolate	[ˈtʃɑːkələt] → [ˈtʃo̞ˈko̞ˌleɪt]	phonemic + nonreduction of stress
basketball	[ˈbæskətˌbɔl] → [ˈbæsukɛˌto̞ˈbo̞ːlu]	anaptyxis
orange	[ˈɔrɪnʤ] → [ɔrˈɪnʤ]	stress placement
nurse	[nɜrs] → [nɑːs]	phonemic
telephone	[ˈtɛləˌfoʊn] → [ˈtɛlɛˌhoʊn]	phonemic
camera	[ˈkæmərə] → [ˈkæˈmɛˌrɑ]	phonemic + nonreduction of stress
doctor	[ˈdɑktər] → [ˈdo̞kuˌtɑː]	anaptyxis, stress
seven	[ˈsɛvən] → [ˈsɛbun]	phonemic

The test was designed to go from the most open responses to the most controlled. Therefore, the first 10 slides comprised the free-response section. Participants were shown a visual aid, along with a prompt which elicited the target word in the response. For example, when the target word was "basketball", the participants were shown a depiction of kids playing

[1]The full instrument will be made available upon request and on the IRIS Database upon publication (see Marsden, Mackey, & Plonsky, 2016).

basketball with the prompt, 子供たちは何のスポーツをしていますか？ (*What sport are the kids playing?*) with the expectation of, "They're playing basketball" as a probable response.

Next were the controlled-production items in which students were tasked with translating sentences embedded with the target words from Japanese to English. For example, for the target word "milk", the elicitation sentence, 牛乳は飲めません (*Gyuunyuu wa nomemasen*) was provided, to which the correct translation is: "I cannot drink milk". Although this section was one of controlled-production, all efforts were made to use both target words and elicitation sentences that were within the ability of the participants to translate quickly and smoothly to obtain the most spontaneity of speech as possible. For example, only the basic grammar forms of present and past tense were used. Errors in translation were not problematic as long as the target word was present in the response.

Finally, the last 10 slides were a word list with each target word given its own slide. The participants were instructed to read each word out loud, in isolation. The same three-part testing instrument was used for the collection at all three time points: pretest, posttest (immediately after the second treatment), and delayed posttest (exactly two weeks after the second treatment). The potential risk of a "practice effect" was considered, though the presence of the control group would allow us to quantify this effect. If any gains were seen by the CG, this effect size would be subtracted from gains seen by the experimental groups, thus adjusting for any practice effect gains, if present.

2.3. Procedures

Each experimental group received two treatment sessions of 30 minutes in duration from the lead researcher, with all groups following the same pattern of: explicit instruction lecture (10 minutes), teacher-led activities (10 minutes), pair work with a classmate (5 minutes) and finally worksheet completion (5 minutes).

2.3.1. Syllabic perception instruction (SPe) group.

Both syllabic experimental groups received explicit instruction on the nature and importance of syllables in English, along with how the concept of syllables varies with that of the Japanese *mora* (for a description of mora, see Nakashima, 2006). Perception-based pedagogy were created following Gilbert's (2008) assertion that, "all practice with rhythm, including the number of syllables, should be accompanied by some physical gesture, such as tapping the desk or moving a foot. This physical movement is far more effective than merely taking mental notice" (p. 37). The SPe group practiced the identification and counting of syllables with the use of hand clapping. Claps were performed to indicate position and relative strength of syllables in words or phrases that the instructor produced. For instance, the instructor would say the word "basketball", to which the participants should clap their hands in the pattern, *CLAP-clap-clap,* indicating three syllables, and stronger on the first one to indicate stress. Where errors were made, the instructor would give corrective feedback by way of repeating the targets with an overemphasis on the stress and syllabic breakdown, or by demonstrating the proper clapping rhythm.

Initially, the 10 items used on the assessment tests were used for illustrations. The target words were shown on the projector screen, and students were asked to clap out the syllables without any modelling from the instructor. Then, the instructor would say the words in a typical Japanese accent, followed by an American accent, to allow the students to attempt to hear the difference in syllabic structure. This side-by-side comparison was accompanied by explicit instruction on features specific to each language, such as how Japanese morae are all relatively equal in length and stress, whereas English syllables are stressed, unstressed, and sometimes deleted altogether (elision). The instructor would continue using the comparison of a Japanese accent versus an American one to illustrate the differences in both single words and entire sentences in order to heighten the students' awareness of the fundamental differences and encourage noticing. Once class-wide practice had achieved relative accuracy, individual students were singled out during drills and made to clap out words and phrases posed by the instructor.

Upon completion of the hand-clapping exercises, a "Syllable Recognition Task" worksheet (see Appendix A) of 20 words was provided. The instructor would read the items aloud and the students were tasked with (a) drawing slashes to separate the words into syllables, and (b) denoting stress by the addition of an apostrophe above the stressed syllable in each word. Students were allowed to work in pairs and the modellings were repeated three times each, allowing the students to discuss, formulate a hypothesis, and then test it over the three repetitions. The answers were discussed with the instructor exaggerating the pronunciations until the students could accurately perceive the breaks and stresses in the items. This activity was not graded by the instructor, but students were asked to self-correct their worksheets in red ink. These worksheets were collected and examined to see if any trends could be detected, the product of which will follow in the Discussion.

2.3.2. Syllabic production instruction (SPr) group.

As previously stated, this group also received explicit instruction on the nature and importance of syllables in English, along with how the concept of syllables varies with that of Japanese *mora*. However, as a production-based treatment group, the SPr focused on the production of words and phrases by closely mimicking the instructor. The instructor did not provide examples of the typical Japanese accent, rather, an American accent was modelled for the students to imitate, with any inaccuracies being corrected by recasting at a slower speed or with exaggerated emphasis on the syllabic structure. Recasts were sometimes accompanied by explicit explanations of what mistakes the students were making in their pronunciation, the most typical difficulties being in the non-target-like addition of vowels (anaptyxis) due to transfer from L1 Japanese.

A "Syllable Production Task" worksheet was provided (see Appendix B), essentially the Syllable Recognition Task of the SPe group with the answers already completed. The students were grouped in pairs, and tasked with verbally producing the items, following the syllable breaks and stresses as they were written. After a practice period, the students individually performed each enunciation and the instructor monitored and gave verbal feedback. This activity

was not graded, but any errors in pronunciation were discussed and repeated until the students executed a satisfactory level of pronunciation accuracy.

2.3.3. Phonetic perception instruction (PPe) group.

The phonetic groups were both focused on segmental aspects of English pronunciation, with no attention being explicitly drawn to the suprasegmental features. The PPe group was given explicit instruction on the different phonemes that were applicable for the pronunciation of the items on the assessment. Most of these were minimal pairs, such as /f/ - /h/, /v/ - /b/, but the three vowels /ɑ/, /æ/, and /ʌ/ were introduced together.

The students were given descriptions of the various phonemes, such as tongue placement and lip formation, and given demonstrations of the differences between the minimal pairs. After the explanations, drills were conducted where the two phonemes of a minimal pair were written on either side of the blackboard. The instructor would produce one of the phonemes, either in isolation or as part of a word. The students were tasked with listening to the utterance, deciding which phoneme had been produced, and indicating which one they thought they recognized by raising either their left or right hand, corresponding to the phonemes on the blackboard.

Once a relative degree of accuracy was achieved in this drill, a "Phonetic Recognition Task" worksheet was distributed (see Appendix C), which included 20 questions. Each item had three or four phonemes or words to choose from. The task for the students was to listen to the instructor's model and choose which phoneme or word was being spoken. The targets were modelled three times each. Answers were discussed as a class, with students asked to correct any errors they made in red ink. While this exercise was not graded, a visual inspection of the students' worksheets indicated that students were generally successful at identifying the correct phonemes. These errors were discussed, with the instructor exaggerating the lip or tongue positions while producing the phonemes until the students could recognize the difference between the phonemes.

2.3.4. Phonetic production instruction (PPr) group.

The PPr group was explicitly taught the same phonemes as the PPe group, but they were made to practice verbalizing them in isolation and in words and phrases, imitating the instructor. The minimal pairs were practiced in contrast to each other, along with a third variation, the typical Japanese accent, which is often a hybrid melding of the two phonemes. The students were given a "Phonetic Production Task" (see Appendix D), a 15-question worksheet which contained words and phrases which utilized the target phonemes. Students were tasked with first practicing to produce all the items in pairs, followed by a performance in front of the class for the teacher. The students would deliver an utterance and the teacher would write on the board what he heard and confirm if it was what the student had been trying to say. Any failures in comprehension were discussed and the students were allowed to recast until the instructor could discern the intended phonemes.

2.4. Scoring

In keeping with current standards in assessing the effects of pronunciation instruction, we adopted expert coders' (i.e., raters with linguistic or teaching backgrounds) subjective analyses

of pronunciation accuracy (Isaacs & Thomson, 2013; Saito & Plonsky, 2019). Each utterance of the target words was rated on a 9-point Likert scale, where only the ends of the scale were defined. A 9-point Likert scale has been suggested to be the ideal range, as smaller scales run the risk of hitting a ceiling effect (Southwood & Flege, 1999). As this study focused on pronunciation *accuracy*, a score of "1" was defined as being "utterly inaccurate" and a score of "9" being "perfectly accurate".

In preparation for data analysis, the lead researcher ran all the recorded voice data through the editing software, AVS Audio Editor v.7. Clips were trimmed to cut out distractions or irritants (such as extended periods of silence), volume levels were normalized, and noise reduction was run where it was deemed necessary.

The first author (a bilingual speaker of American English/Japanese, with a background in applied linguistics and over 20 years teaching experience in Japan) was the main expert coder. In order to demonstrate support for the consistency and validity of the rating procedure, two additional expert coders were asked to assist in the ratings. Rater 2 was a bilingual American English/Japanese speaker with a background in applied linguistics and six years teaching experience in Japan. Rater 3 was an L1 speaker of Australian English, with lower intermediate Japanese L2 ability, and had approximately one year of teaching experience in Japan. The three raters participated in a 30-minute training session in which the rating system was explained, and actual recordings from the piloting sessions were reviewed together and discussed. As the additional raters' data were only necessary for establishing the validity of the first author's pronunciation accuracy ratings, only a subset of 1/3 of the total set of recordings was rated by Raters 2 and 3.

Once the scoring was complete, Pearson's correlation coefficients (r) were run on the portion of data which all three raters had assessed (Table 2). The correlations between all raters were positive and fairly strong, especially in the case of Rater 1 and Rater 2 (r = .78; both L1 American English speakers). Rater 3 (r = .56 with Rater 1; r = .57 with Rater 2) indicated in debriefing that she was generally more lenient in her scoring due to her familiarity with both American and Australian pronunciation. Although perfect agreement was not observed, the interrater reliability was interpreted as indicating that others scoring the same data would arrive at largely similar conclusions about the accuracy of the participants' production, particularly in the framework of American English (see Plonsky & Derrick, 2016; Saito & Plonsky, 2019). For the remainder of the discussion of the results and statistical analyses, only the full dataset produced by the first author (who not only had the most teaching experience, but also conducted the experimental treatments for all groups) was used.

Table 2

Pearson Correlation Coefficients for Overall Test Ratings

	Rater 1	Rater 2	Rater 3
Rater 1	-	.78**	.56**
Rater 2	-	-	.57**
Rater 3	-	-	-

** Correlation is significant at the 0.01 level (2-tailed)

2.5. Analysis

In order to answer the research question (*What are the relative effects of perception-based versus production-based training of syllables and phonemes on pronunciation accuracy?*), we first calculated gain scores for each condition from the pretest to the immediate posttest and to the delayed posttest. Descriptive statistics at each time point, as well as for these two sets of gain scores, were then examined and compared formally using ANOVA. Effect sizes expressing the standardized mean difference (*d*) for all contrasts were also calculated to shed light on the magnitude of group differences.

Prior to these analyses, however, the statistical assumptions for ANOVA were checked. Normality was assessed visually (i.e., by examining histograms, boxplots, and Q-Q plots for each group) as well as more formally by means of the Shapiro-Wilk test. All indicators pointed to a reasonably normal distribution both for the pre-post and pre-delayed posttest gain scores. There was less evidence for homogeneity of variance across groups, however, with a narrower spread of scores found for the control group. We do not believe that this slight violation of the assumptions has had a substantial impact on our results. Nevertheless, our findings should be interpreted with this statistical artifact in mind.

3. Results

The study showed that while all experimental groups experienced significant gains in their pronunciation accuracy as a result of explicit instruction, perception-based instruction clearly resulted in larger gains, particularly over time. Tables 3 and 4 present the descriptive statistics at each testing point as well as for change (gain) scores across all groups. The control group's gains were negligible (.13 and .15, respectively) with corresponding confidence intervals that touch or nearly touch zero. We can therefore dismiss the likelihood of practice effects. By contrast, gains can be observed for all four treatment groups on both the immediate and delayed posttests relative to the pretest. In fact, three of the four treatment groups appear to have continued an upward trajectory beyond the immediate posttests (SPe, PPe, and PPr), although the increase was quite small in the case of PPr.

Table 3

Descriptive Statistics for Pretest, Posttest, and Delayed Posttest Scores

	Pretest		Posttest		Delayed Posttest	
Group	*M* (*SD*)	95% *CI*	*M* (*SD*)	95% *CI*	*M* (*SD*)	95% *CI*
CG	3.92 (.34)	[3.77, 4.06]	4.05 (.46)	[3.86, 4.25]	4.07 (.45)	[3.87, 4.26]
SPe	3.92 (.48)	[3.70, 4.13]	5.42 (.64)	[5.13, 5.70]	5.64 (.64)	[5.35, 5.93]
SPr	4.27 (.43)	[4.08, 4.46]	5.63 (.37)	[5.47, 5.79]	5.25 (.50)	[5.02, 5.47]
PPe	3.83 (.38)	[3.66, 3.99]	4.42 (.76)	[4.10, 4.74]	5.06 (.55)	[4.83, 5.29]
PPr	4.23 (.48)	[4.03, 4.42]	5.35 (.86)	[5.00, 5.69]	5.38 (.72)	[5.08, 5.68]

Note. CG = control group; SPe = syllabic perception; SPr = syllabic production; PPe = phonemic perception; PPr = phonemic production

Table 4

Descriptive Statistics for Gain Scores

	Pretest to Posttest		Pretest to Delayed Posttest	
Group	*M* (*SD*)	95% *CI*	*M* (*SD*)	95% *CI*
CG	0.13 (.31)	[0.00, 0.27]	0.15 (.30)	[0.02, 0.28]
SPe	1.50 (.63)	[1.22, 1.79]	1.72 (.64)	[1.44, 2.02]
SPr	1.36 (.34)	[1.21, 1.51]	0.98 (.52)	[0.74, 1.21]
PPe	0.59 (.72)	[0.29, 0.89]	1.24 (.51)	[1.02, 1.45]
PPr	1.12 (.86)	[0.76, 1.47]	1.15 (.66)	[0.88, 1.42]

Note. CG = control group; SPe = syllabic perception; SPr = syllabic production; PPe = phonemic perception; PPr = phonemic production

In order to more formally address our research question concerning the relative effects of different types of PI over time, we conducted two analyses of variance to compare group gain scores from the pretest to the immediate posttest and from the pretest to the delayed posttest. The differences across groups from pre to posttest was found to be statistically significant with group membership explaining 41% of the variance in change scores ($F = 19.04$; $df = 4, 110$; $p < .001$; $\eta^2 = .41$). The relative size of the gains across the five groups is expressed as *d* values in Table 5. The differences in gains between both syllable groups (SPe, SPr) and the control group are very large ($d = 2.79$ and 3.76, respectively). The relative difference in gains for the two groups that received phonemic instruction are large as well ($d = .82$ for PPe and 1.49 for PPr) but were substantially smaller by comparison.

Table 5

*Effect sizes (*d*) for Contrasts Across Pre-Post Gain Scores*

Group	CG	SPe	SPr	PPe	PPr
CG	-	-	-	-	-
SPe	2.79	-	-	-	-
SPr	3.76	-0.29	-	-	-
PPe	0.82	-1.35	-1.35	-	-
PPr	1.49	-0.51	-0.36	0.66	

Note: *d* values here refer to standardized mean differences. The group on the horizonal axis is the 'reference' group. Positive values can therefore be understood as indicating a higher score for the group in the far-left column.

The differences across groups from pre to delayed posttest was statistically significant as well ($F = 25.33$; $df = 4, 110$; $p < .001$). In this case, treatment condition explained nearly half of the variance across groups ($\eta^2 = .48$). The relative size of the gains across the five groups is expressed as *d* values in Table 6. The differences in gains between all four treatment and the control group are very large (SPe, SPr, PPe, and PPr obtained *d* values of 3.22, 1.95, 2.67, and 1.94, respectively). For this set of gain scores, however, the largest gains appear to be in favor of the two perception-based conditions, SPe and PPe (3.22 and 2.67).

Table 6

*Effect sizes (*d*) for Contrasts Across Pre-Delayed Posttest Gain Scores*

Group	CG	SPe	SPr	PPe	PPr
CG	-	-	-	-	-
SPe	3.22	-	-	-	-
SPr	1.95	-1.29	-	-	-
PPe	2.67	-0.86	0.51	-	-
PPr	1.94	-0.89	0.29	.15	-

Note: *d* values here refer to standardized mean differences. The group on the horizonal axis is the 'reference' group. Positive values can therefore be understood as indicating a higher score for the group in the far-left column.

4. Discussion

This study sought to determine what the relative effects of perception-based versus production-based training of syllables and phonemes would be on L1 Japanese tertiary students' pronunciation accuracy. Based on current standards in measuring effect sizes for L2 field-specific research (Plonsky & Oswald, 2014), the results showed first, that all treatments resulted in significant and large gains in participants' ability to accurately produce the target language. The change observed generally aligns with Lee et al.'s (2015) and Saito and Plonsky's (2019) findings regarding the positive effects of pronunciation instruction. Furthermore, our study also demonstrates the durability of such effects as shown on delayed posttest measures. These

findings are significant, as they add further evidence to the notion that pronunciation instruction is indeed effective, and that its effects hold up across different instructional targets and techniques (Saito, 2012; Thomson & Derwing, 2015).

However, the current study was not solely concerned with testing the general effects of PI. We were particularly interested in comparing gains across four different types of instructional conditions that varied in terms of being syllable- vs. phoneme-based and perception- vs. production-based. When comparing the gains made by each of the four groups, the results seem to indicate an advantage in immediate gains following syllable-based (over phoneme-based) instruction. Regarding the question of whether instruction that involves production is necessary to improve learners' productive performance, the answer appears to be 'no' (see Shintani et al., 2013). Whereas immediate gains showed an advantage for the condition that received production-based practice, the opposite pattern is evident from the results on the delayed posttest. Differences in effect size of between .73 and 1.27 were observed in each pairing, allowing this perception-based advantage to be interpreted as medium-to-large according to L2 field-specific research standards (Plonsky & Oswald, 2014). These results fell in line with our hypothesis that perception-based instruction would lead to greater improvement than production-based ones, for both features.

While Flege's SLM mainly addresses the process by which learners must first be able to perceive foreign language phonemes before they are able to produce them, this study suggests that this model may generalize to suprasegmental features as well. The worksheets completed by each group, though not used in the final data analysis, did reveal some noteworthy patterns. The most revealing group was the syllabic perception group (SPe). A visual inspection of their worksheets showed that most students were still unsure as to what constituted a syllable but were nevertheless quite successful in picking up the (suprasegmental) stress patterns of the items. For shorter items with two or three syllables, detection of stress equated to the proper syllable divisions. However, for longer items, syllables were often under-detected. As an illustration, common mistakes were dividing the words *university* and *watermelon* as [uni / vér / sity] and [wáter / melon], instead of [u / ni / vér / si / ty], and [wá / ter / me / lon].

Although still technically incorrect, this represents a significant shift from the pretest conditions, where typically L1 Japanese speakers tend to over-detect the number of syllables in English words due to the Japanese phonetic C-V rule. A word such as *Christmas*, which has only two syllables in English, is usually pronounced as a 5-syllable word, [ku / ri / su / ma / su], when spoken by an untrained L1 Japanese speaker. The accurate detection and subsequent replication of suprasegmental stress patterns (increased blending of sounds) by the participants most likely played a major role in the increased rating of SPe members' post- and delayed posttest pronunciation accuracy, as compared with the other groups. This is a significant finding that directly lends support to the SLM, which posits that accurate perception precedes accurate production in the L2. It should also be noted that the two perception-based groups (SPe and PPe) were the two groups to show not only the greatest degrees of improvement overall, but also in the gap between post- and delayed posttest. This finding further extends those of Sakai and

Moorman's (2018) showing strong gains for perception-based instruction at the meta-analytic level.

4.1. Pedagogical Implications

Beyond testing the SLM, the present study was also interested in testing pedagogical models for teachers looking to incorporate PI into their curriculum. From our results, we can recommend perception-based methodologies such as kinesthetic training or phoneme identification tasks take priority, especially if classroom time dedicated to PI is limited. These results also align with other studies which have suggested that hand-clapping or rhythmic beating benefit L2 pronunciation development (Gluhareva & Prieto, 2017; Zhang & Baills, 2018). It should be noted that these techniques are not solely restricted to Japanese learners or tertiary education contexts but can be thought of as universally applicable. While this study only made use of mimicking exercises such as tapping or clapping to the beat, there are plenty of other techniques that can be created based on this principle to match specific circumstances or age groups. For example, younger learners may enjoy hopping in time with the beat, smaller jumps for unstressed, and bigger jumps for stressed syllables. In situations where a quieter method is preferable, stretching a rubber band between the hands will allow the learners to mimic the timing and stress of syllables by pulling it longer or shorter (see Gilbert, 2008 for a description of this technique). Burri and Baker (2016) also assert that learning pronunciation can be fun and easy by using a perception-based kinesthetic/haptic system called the Butterfly Technique. This technique differs slightly from other kinesthetic systems in that students use separate hands to tap different places on their bodies when identifying strong or weak syllables (e.g., tapping their right shoulder strongly with their left hand, and their left elbow lightly with their right hand.)

Of course, implementing PI in an actual classroom would not need to artificially restrict students to only perception *or* production as in the present, lab-based study. Since all four conditions reported here were shown to have positive effects on improving pronunciation accuracy, the case could be made for implementing a much more realistic, blended approach which uses elements of both to maximize effects. To illustrate this point, students in the perception groups (SPe and PPe) were discouraged from repeating the model utterances from the instructor to keep the treatment as production-free as possible. However, such a restriction is impractical and perhaps counterproductive. In reality, it is natural for students to repeat the instructor's models (engage in production) in order to test their assumptions and abilities, which in effect trains perception and production simultaneously. It is becoming to be accepted that both preceptive and productive dimensions are interrelated and complementary (Nagle, 2018a, 2018b), thus supporting the incorporation of both modes into a methodology concurrently.

4.2. Limitations and Directions for Future Research

This study attempted to assess the immediate and delayed effects of four different PI treatment methods in relation to each other and to a control group. Although we are confident that our sample was sufficient to detect a relatively stable effect, the groups were fairly small. Future studies could further add statistical robustness to our findings by reducing the number of experimental groups and instead focusing on larger populations of each.

While it would normally be expected that gains observed at the immediate posttest would gradually decrease in the absence of further training, gains for the present study extended beyond the treatment period in all groups except SPr. It is not clear why this was the only group to give up some of its gains, but even at d = 1.95 at the delayed posttest, this is a significant increase over pretest conditions. As the treatments only lasted two weeks, it is likely that the duration was not long enough for learners to fully form phonetic representations and they were still relying heavily on memory. Future studies may wish to utilize a longer treatment duration.

Finally, the assessed segmental range of this study only included 10 phonemes, found to be problematic for Japanese learners of English in the piloting phase: /a/, /ɑ/, /ɜː/, /o/, /f/, /h/, /v/, /b/, /s/ and /θ/ (see Saito, 2015; Gooch, Saito, & Lyster, 2016; for separate research on Japanese /r/ vs /l/). On the suprasegmental side, features included syllables, stress, and the lax syllable /ə/. This set of features requires us to interpret results tentatively as it leaves large gaps in the research regarding the relative effects of PI on other target features in English. Further research is therefore urgently needed which targets other segmental phonemes, suprasegmental features, L1 populations, and even other L2 target languages to establish the generalizability of our findings.

5. Conclusion

On a theoretical front, scholars have thus far explored whether L2 speech learning is perception- or production-based (e.g., Flege, 1995 for SLM vs. Best and Tyler, 2007 for PAM-L2), and whether such tenets could be generalizable across both segmental and suprasegmental dimensions (e.g., Trofimovich & Baker, 2006; So & Best, 2010). To examine the generalizability of such theoretical discussion to the context of English-as-a-Foreign-Language classrooms, the current study examined the relative effectiveness of perception- and production-based training on segmental and suprasegmental aspects of L2 pronunciation development by 115 Japanese college-level students. The results suggested that perception-based training is more effective than production-based training; and that such results can be generalized across different dimensions of L2 speech learning (segmentals vs. suprasegmentals).

6. References

Aliaga-Garcia, C. & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In: M. A. Watkins & B. O. A. Baptista (Eds.) *New Sounds 2007: Proceedings of the 5th International Symposium on the Acquisition of Second Language Speech* (10-27). Florianópolis, Brazil: Federal University of Santa Catarina.

Baker, W. & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics in Language Teaching, 44*(3), 231-250.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In: J. Goodman & H. C. Nusbaum (Eds.) *The Development of Speech Perception: The Transition From Speech Sounds to Spoken Words* (167-224). Cambridge, MA: MIT Press.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In: W. Strange (Ed.) *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research* (171-206). Timonium, MD: York Press Inc.

Best, C. T. & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In: M. J. Munro & O. S. Bohn (Eds.) *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (13-34). Amsterdam: John Benjamins.

Bradlow, A., Pisoni, D., Akahane-Yamada, R. & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Acoustical Society of America, 101*(4), 2299-2310.

Burri, M. & Baker, A. A. (2016). Teaching rhythm and rhythm grouping: The butterfly technique. *English Australia Journal: The Australian Journal of English Language Teaching, 31*(2), 72-77.

Derwing, T. M. & Munro, M. J. (2015). *Pronunciation Fundamentals: Evidence-Based Perspectives for L2 Teaching and Research*. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning, 27*(2), 315-330.

Eckman, F. R. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition, 13*(1), 23-41.

Ellis, R. (1997). *Second Language Acquisition.* Oxford, UK: Oxford University Press.

Ellis, R. (2001). Investigating form-focused instruction. *Language Learning, 51*(s1), 1-46.

Ellis, R. (2003). *Task-Based Language Learning and Teaching.* Oxford, UK: Oxford University Press.

Flege, J. E. (1988). The production and perception of foreign language speech sounds. In: H. Winitz (Ed.) *Human Communication and its Disorders, A Review* (224-401). Norwood, NJ: Ablex.

Flege, J. E. (1992). Speech learning in a second language. In: C. Ferguson, L. Menn & C. Stoel-Gammon (Eds.) *Phonological Development: Models, Research, and Implications* (565-604). Timonium, MD: York Press.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America, 93*(3), 1589-1608.

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* (229-273). Timonium, MD: York Press.

Flege, J. E. (1999). Age of learning and second-language speech. In: E. Birdsong (Ed.) *New Perspectives on the Critical Period Hypothesis for Second Language Acquisition* (101-132). Hillsdale, NJ: Lawrence Erlbaum.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In: P. Burmeister, T. Piske, & A. Rohde (Eds.) *An Integrated View of Language Development: Papers in Honor of Henning Wode* (217-244). Trier: Wissenschaftlicher Verlag.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In: J. Cole & J. I. Hualde (Eds.) *Laboratory Phonology 9* (353-381). Berlin/New York: Mouton de Gruyter.

Flege, J. E., Bohn, O-S & Jang, S. (1997). The effect of experience on non-native subjects' production and perception of English vowels. *Journal of Phonetics, 25*(4), 437-470.

Gilbert, J. B. (2008). *Teaching pronunciation using the prosody pyramid.* Cambridge: Cambridge University Press.

Gluhareva, D. & Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research, 21*(5), 609-631. http://dx.doi.org/10.1177/1362168816651463

Gooch, R., Saito, K., & Lyster, R. (2016). Effects of recasts and prompts on L2 pronunciation - development: Teaching English /r/ to Korean adult EFL learners. *System, 60*, 117-127. https://doi.org/10.1016/j.system.2016.06.007

Heinz, M. (2013). A preliminary survey of the preferred learning methods for interpretation students. *Journal of International Education Research, 9*(4), 293-304. http://dx.doi.org/10.19030/jier.v9i4.8081

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly, 10*, 135-159. https://doi.org/10.1080/15434303.2013.769545

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics, 36*(3), 345-366. http://dx.doi.org/10.1093/applin/amu040

Levis, J. (2018). *Intelligibility, Oral Communication, and the Teaching of Pronunciation.* Cambridge: Cambridge University Press.

Lyster, R. (2007). *Learning and Teaching Languages Through Context: A Counterbalanced*

Approach. Amsterdam / Philadelphia: John Benjamins.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching, 46*(1), 1-40. http://dx.doi.org/10.1017/S0261444812000365

Marsden, E., Mackey, A., & Plonsky, L. (2016). Breadth and depth: The IRIS repository. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS repository of instruments for research into second languages* (pp. 1-21). New York: Routledge.

Mart, C. T. (2013). The Grammar-Translation Method and the use of translation to facilitate learning in ESL classes. *Journal of Advances in English Language Teaching, 1*(4), 103-105.

Major, R. C. (1986). The ontogeny model: Evidence from L2 acquisition of Spanish r. *Language Learning, 36*(4), 453-504.

Major, R. C. (2001). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology.* Mahwah, NJ: Lawrence Erlbaum Associates.

Munro, M. J. (1993). Production of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech, 36*(1), 39-66.

Nagle, C. (2018a). Perception, production, and perception-production: Research findings and implications for language pedagogy. *Contact, 44*(2), 5-12.

Nagle, C. L. (2018b). Examining the temporal structure of the perception-production link in second language acquisition: A longitudinal study. *Language Learning, 68*, 234-270. https://doi.org/10.1111/land.12275.

Nakashima, T. (2006). Intelligibility, suprasegmentals, and L2 pronunciation instruction for EFL Japanese learners. *福岡教育大学紀要 (Bulletin of Fukuoka University of Education), 55*(1), 27-42.

Plonsky, L. & Derrick, D. J. (2016). A meta-analysis of reliability coefficients in second language research. *Modern Language Journal, 100*(2), 538-553. http://dx.doi.org/10.1111/modl.12335

Plonsky, L. & Oswald, F. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878-912. http://dx.doi.org/10.1111/lang.12079

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly, 46*(4), 842-854. https://doi.org/10.1002/tesq.67

Saito, K. (2015). Communicative focus on second language phonetic form: Teaching Japanese learners to perceive and produce English /ɹ/ without explicit instruction. *Applied Psycholinguistics, 36*(2), 377-409. https://doi.org/10.1017/S0142716413000271

Saito, K. & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed framework and meta-analysis. *Language Learning, 69*(3), 652-708. http://dx.doi.org/10.1111/lang.12345

Saito, K. & Wu, X. (2014). Communicative focus on form and second language suprasegmental

learning: Teaching Cantonese learners to perceive Mandarin tones. *Studies in Second Language Acquisition, 36*(4), 647-680. http://dx.doi.org/10.1017/S0272263114000114

Sakai, M. & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics, 39*(1), 187-224.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158.

Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-based versus production-based grammar instruction: A meta-analysis of comparative studies. *Language Learning, 63*(2), 296–329.

So, C. & Best, C. (2010). Perception of initial obstruent voicing is influenced by gestural organisation. *Journal of Phonetics, 38*(2010), 109-126.

Southwood, M. H. & Flege, J. E. (1999). Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics & Phonetics, 13*(5), 335-349.

Thomson, R. I. & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics, 36*, 326-344. https://doi.org/10.1093/applin/amu076

Trofimovich, P. & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28*(01), 1-30.

VanPatten, B. (2007). Input processing in adult second language acquisition. In: B. VanPatten & J. Williams (Eds.) *Theories, Research, and Commentary* (5-32). Mahwah, NJ: Erlbaum.

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*, 1-24. http://dx.doi.org/10.1177/1362168818806531

7. Appendices

Appendix A

Syllable Recognition Task / 音節クイズ

Break up the words into their syllable components by drawing a slash (/) at each break. Also, put an accent mark (‘) on the syllable that takes the main stress. 以下の英単語をスラッシュ（/）で音節ごとに分けて下さい。そして、一番強い音節の上に（‘）を付けて下さい。（２回目：先生の声を聞きながらもう一度赤ペンでやって下さい。）

1) b a s k e t b a l l
2) w a t e r m e l o n
3) c o m p u t e r
4) E n g l i s h
5) m i l k
6) c h o c o l a t e
7) c a k e
8) s o c c e r
9) u n i v e r s i t y
10) f i r e f i g h t e r
11) c a m e r a
12) s t a t i o n
13) t e l e p h o n e
14) D V D
15) e l e v a t o r
16) b l a c k b o a r d
17) c h a l k
18) b a c k p a c k
19) w o r l d
20) J a p a n e s e

Appendix B

Syllable Production Task / 音節クイズ

Practice saying the following words by looking at the syllable patterns. Pay attention to the stressed syllables (in **bold**). Some words have "dropped" syllables (in *italics*). 以下の英単語を音節を注意しながら発言しましょう。**太字** は強めで、*イタリック体* は音をしませんのでご注意ください。

1) **b á s** / k e t / b a l l
2) **w á** / t e r / m e / l o n
3) c o m / **p ú** / t e r
4) **É n** / g l i s h
5) **m í l k**
6) **c h ó c** / *o* l a t e
7) **c á k e**
8) **s ó** / c c e r
9) u / n i / **v é r** / s i / t y
10) **f í** / r e / **f í g h** / t e r
11) **c á m** / *e* r a
12) **s t á** / t i o n
13) **t é** / l e / p h o n e
14) **D** / **V** / **D**
15) **é** / l e / v a / t o r
16) **b l á c k** / b o a r d
17) **c h á l k**
18) **b á c k** / p a c k
19) **w ó r l d**
20) J a / p a / **n é s e**

Appendix C

Phonetic Recognition Task / 音素クイズ

Listen to the teacher and choose which sound (or word) is being said. 先生の声をよく聴いて、どっちの音素や言葉が言われているか選んで、丸を付けて下さい。

1) f / h / v / b
2) f / h / v / b
3) f / h / v / b
4) f / h / v / b
5) s / sh / th / z
6) s / sh / th / z
7) s / sh / th / z
8) s / sh / th / z
9) æ / a / ʌ
10) æ / a / ʌ
11) fay / hay / vay / bay
12) few / hue / view / bue
13) ferry / hery / very / berry
14) fox / hox / vox / box
15) sue / shoe / thue / zoo
16) sa / sha / the / za
17) lesser / lesher / leather / leza
18) tam / Tom / tum
19) bat / bot / butt
20) slat / slot / slut

Appendix D

Phonetic Production Task / 音素クイズ

Practice saying the following mixed pairs. 以下の英単語の発音を練習しましょう。

1) fay / hay / vay / bay
2) few / hue / view / bue
3) ferry / hery / very / berry
4) fox / hox / vox / box
5) sue / shoe / thue / zoo
6) sa / sha / the / za
7) lesser / lesher / leather / leza
8) tam / Tom / tum
9) bat / bot / butt
10) slat / slot / slut
11) cam / calm / come
12) hat / hot / hut
13) That last act was rather bad.
14) The garage is opposite the hock shop.
15) Rusty loves buttered buns

Lee, B., Plonsky, L., & Saito, K. (2019). The effects of perception- vs. production-based pronunciation instruction. *System*. https://doi.org/10.1016/j.system.2019.102185

Is it better to learn how to 'hear' the sounds of a new language, or practice saying them?

What this research was about and why it is important

Pronunciation is one the hardest parts about learning a new language, especially if it is very different from your native one. Each language or dialect has their own set of unique sounds and intonation patterns. So, what is the best way to go about learning the new sounds of a foreign language? At first, it may seem natural that to improve your pronunciation, you should just practice speaking. However, previous research has suggested that your ears have as much to do with pronunciation as your voice does. This study compared four groups of students, trained in different ways and on different targets of pronunciation, to see which combination was the best. Our results suggested that students trained to listen carefully and identify English syllables had the biggest improvement in their own pronunciation.

What the researchers did

- Learners were 115 Japanese university students, aged between 18-20. Students in Japan learn English from 7th to 12th grade, so each student had at least 6 years of English training before the start of the study.
- A list of 10 English words which are hard for Japanese speakers to pronounce were chosen for the test. The students were asked a series of questions, made to translate Japanese into English, and read off a word list, which resulted in them saying each of the 10 words three times each. The students didn't know which words they were saying were being tested until the word list, which was at the very end of the test.
- Each group had two special pronunciation lessons, for 30 minutes each, in one of the following styles: listen and identify English sounds (e.g., did I say /f/ or /v/ or /b/?); listen and identify English syllables (e.g., clap your hands for each syllable you hear); listen and repeat sounds (e.g., lightly bite down on your bottom lip and push air out to say the sound /f/); listen and repeat focusing on syllables.
- The test was repeated after the two lessons were finished, and again after a gap of two weeks.
- Three English professors (native speakers) individually listened to the test recordings and scored each students' pronunciation. All three professors' scores generally agreed with each other.

What the researchers found

- All groups greatly improved their pronunciation as a result of the two lessons.
- At first, the group which was practicing listening and repeating English syllables improved the most. However, after a two-week gap of having no lessons, this group's scores dropped a bit while the three other groups continued to climb.
- By the end of the study, the group which practiced listening and identifying the correct number of syllables had the most improvement in their own pronunciation, followed by the group which practiced listening and identifying correct sounds.

Things to consider

- One of the biggest takeaways of this study was that all the students improved as a result of the special lessons. Teachers should be encouraged to add pronunciation instruction to their lessons wherever possible.
- For teachers who would like to start teaching pronunciation but are not quite sure how, this study offered four techniques that were all shown to be effective, in addition to providing a discussion of other possible ideas and links to other researchers' studies.
- This study helped add evidence to support the hypothesis that training in listening skills can carry over to help improve students' pronunciation as well.
- A unique aspect of the study was that it compared listening vs. speaking and sounds vs. intonation at the same time, using multiple groups. A key finding is that both listening groups improved more than both speaking groups, and that they continued to get better even after the lessons ended. This result should be followed up on in further research.
- As this study only looked at 115 Japanese university students, much more research is needed. Future directions could include different language pairs, younger (or older) students, different sounds, etc.

Materials available from https://www.iris-database.org/
How to cite this summary: Lee, B., Plonsky, L., Saito, K. (2019). Is it better to learn how to 'hear' the sounds of a new language, or practice saying them? *OASIS Summary* of Lee, Plonsky, & Saito (2019) in *System* https://oasis-database.org