The Perception-Production Link Revisited: The Case of Japanese Learners' English /ɹ/ Performance

Kazuya Saito & Kim van Poeteren

Abstract (English)

The current study re-examines how second language speech perception and production are related in the context of the acquisition of English /ɹ/ by 45 adult Japanese learners with various proficiency levels. Perception was evaluated using a two-alternative forced choice identification task, while pronunciation performance was assessed via multiple task/analysis contexts. Overall, the participants' perception performance was correlated with the global qualities (accuracy, intelligibility) of their production ability both at controlled and spontaneous speech levels. In light of the results of acoustic analyses, however, their perception ability was only significantly predictive of their redeployment of existing articulatory parameters (i.e., lower F2 for the rate and degree of tongue retraction), not the acquisition of new articulatory parameters (i.e., lower F3 for labial, palatal, and pharyngeal constrictions).

Key words: Second language phonetics, English /ɹ/, pronunciation, speech production, speech perception

Abstract (Dutch)

Deze studie herexamineert hoe de perceptie en productie van een tweede taal verbonden zijn in het kader van de verwerving van het Engels /ɹ/ bij 45 Japanse volwassenen met verschillende vaardigheidsniveaus. Perceptie werd geëvalueerd door gebruik van een luistertaak, terwijl uitspraak prestatie werd vastgesteld via meerdere taak/analyse contexten. In het algemeen werden de prestaties van de deelnemers op vlak van perceptie gecorreleerd met de globale eigenschappen (nauwkeurigheid, begrijpelijkheid) van hun voorbrengingsvermogen zowel aan gecontroleerde en spontane spraakniveaus. Volgens de resultaten van akoestische analyses was hun perceptievermogen alleen aanzienlijk voorspellend van hun herinzetten van bestaande akoestische parameters (d.w.z., lagere F2), en niet van de verwerving van nieuwe akoestische parameters (d.w.z., lagere F3).

Key words: Tweede taal fonetiek, Engels /ɹ/, uitspraak, spraakproductie, spraakperceptie

The Perception-Production Link Revisited: The Case of Japanese Learners' English /ɹ/ Performance

Whereas a number of second language (L2) phonetics studies have examined how adult L2 learners can improve their ability to perceive and produce new sounds though increased experience and interaction with native and non-native speakers, there is a general consensus that such perception and production domains are somewhat interrelated (e.g., Flege, 2003 for a perception-first view; Best & Tyler, 2007 for a direct-realist position). At the same time, it is noteworthy that adult L2 speech performance has been identified as a multifaceted phenomenon according to various influencing factors, such as learners' relevant experience (inexperienced vs. experienced: Trofimovich & Baker, 2006), target sound categories (phonetic features similar and dissimilar to L1 counterparts: McAllister, Flege, & Piske, 2002), and task conditions (controlled vs. spontaneous: Major, 2008). In the context of English /ɹ/ performance by 45 native speakers of Japanese (NJs) with various proficiency levels, the current study aimed to scrutinize the complex relationship between L2 perception and production ability from various angles. In particular, the study investigated the predictive role of L2 perception ability for psycholinguistic (native judges' overall accuracy/intelligibility assessment) and acoustic (F3, F2, F1 and duration) aspects of L2 production ability at controlled and spontaneous speech levels.

Background

Second Language Speech Perception and Production

In the L2 speech learning literature, there exists much discussion on the relationship and directionality of perception and production. One of the most influential L2 speech theories related to the interaction between perception, production, and development is the Speech Learning Model developed by Flege and colleagues (Flege, 2003; McAllister et al., 2002; Piske, MacKay, & Flege, 2001). The Speech Learning Model states that learners store L2 speech information in the brain based on how they hear new sounds as an acoustic signal. This suggests that the perception of new L2 sounds activates relevant sensorimotor skills and leads to production ability. Indeed, a number of cross-sectional studies have noted not only significant correlations between measures of perception and production for both vowels and consonants, but also the tendency for L2 learners to have better perception than production ability (e.g., Baker & Trofimovich, 2006; Cardoso, 2011; Flege, 1993; Flege, Bohn, & Jang, 1997; Flege, MacKay, & Meador, 1999).

Using a pre/post test design, Bradlow and colleagues conducted a series of studies to explore the perception-production link. The results demonstrated that NJs who received only intensive perception training on the /ɹ/-/l/ contrast without any explicit articulatory instruction not only enhanced their perception abilities, but also transferred the gains to production both in the short and long term (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997) (for similar behavioural results, see also Hardison, 2003; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005; Saito, 2015; Saito & Wu, 2014; Trofimovich, Lightbown, Halter, & Song, 2009). More recently, Zhang et al. (2009) used magnetoencephalography to examine how intensive perceptual training (on the non-native /ɹ/-/l/ contrast for NJs) enhances neural activities relevant for L2 speech production. The results provided some evidence (i.e., increased activity in the left inferior frontal region) that "phonetic learning may strengthen the perceptual–motor link by recruiting the Broca's area" (p. 237).

Another influential account of the perception-production link can be found in gestural theories of speech perception and production, such as Motor Theory (Liberman & Whalen, 2000) or Direct-Realist Theory (Best & Tyler, 2007).[1] Both positions equally assume that L2 speech information is represented in the brain based on relevant articulatory gestures (i.e., how to use the tongue, lips and jaw to produce new sounds) rather than acoustic signals. In this regard, learners develop perception and production simultaneously with increasing awareness of speech gestures; thus, these domains are closely related to one another (but see Bundgaard-Nielsen, Best, Kroos, & Tyler, 2012).

Interestingly, other studies have also found some "asymmetry" in the connection between perception and production performance. For example, it has been reported that some NJs produced the English /ɹ/ and /l/ contrast more accurately than they perceived it (Goto, 1971; Sheldon & Strange, 1982). Similarly, Bohn and Flege (1997) showed that experienced German learners produced the English vowel /ɛ/-/æ/ contrast, but did not accurately perceive it (see also Zampini & Green, 2001). Sheldon and Strange (1982) claimed that whereas improvement in perception may initially lead to changes in production, production abilities can develop irrespective of perception abilities, especially in the case of advanced L2 learners who tend to have explicit articulatory knowledge of L2 sounds and social pressure to be intelligible in their production.

These previous studies generally support the broad idea that perception and production abilities are essentially related to each other. At the same time, however, it remains open to further investigation the reasons why, under different circumstances, different results can be found. For example, major L2 speech theories have shown that certain phonetic sounds entail much learning difficulty, especially when they do not have an L1 counterpart (Baker et al., 2008). In these difficult L1/L2 parings, which have great phonetic distance, L2 learners' perception abilities may precede their production abilities (Goto, 1971) or vice versa (Bradlow et al., 1997). Alternatively, adult L2 learners' perception abilities could be more strongly linked with the intelligibility of their production rather than with their attainment of nativelike pronunciation forms (Abrahamsson & Hyltenstam, 2009). Finally, the Ontogeny Phylogeny Model (Major, 2008) emphasizes the role of styles, especially in adult L2 pronunciation development: the quality of L2 learners' production tends to be more targetlike when tested with controlled (e.g., word reading) compared to free speech tasks (e.g., picture descriptions), indicating that the predictive power of L2 perception ability may display a great deal of variation in relation to how speech production is elicited (see also Derwing & Rossiter, 2003).

To take into account the intricate mechanisms underlying adult L2 speech learning, whose process and product are well-known to be affected by various learner-internal and external factors, the current study was designed to revisit the perception-production link by looking at how adult Japanese learners with various proficiency levels differentially perceived and produced English /ɹ/ under different task conditions.

---

[1] The two theories differ in regard to the existence of a phonetically-specialized module, however. For Motor Theory, it is claimed that humans are endowed with a phonetic module by which they can have access to one single articulatorily-defined representation in the brain which is responsible for both the perception and production of a sound. For the Direct-Realist Theory, it is proposed that the acoustic signal itself provides listeners with information about the corresponding phonetic gestures.

English /ɹ/

Given that the Japanese phonetic system has neither General American English /ɹ/ nor /l/, Japanese learners tend to show tremendous difficulty acquiring this non-native contrast in perception (Lively, Logan, & Pisoni, 1993) and in production (Larson-Hall, 2006). With respect to perception, English /ɹ/ is different from English /l/, with lower formant frequencies (F3 = 1600-1900 Hz; F2 = 1700-2100 Hz; F1 = 250-550 Hz) and longer phonemic duration (> 50 ms) (Espy-Wilson, 1993; Espy-Wilson et al., 2000; Flege, Takagi, & Mann, 1995; Hattori & Iverson, 2009). Because Japanese learners do not use F3 information as a primary acoustic cue for distinguishing any consonant and vocalic contrasts, they are likely to ignore F3 variation when asked to distinguish between English approximant categories (/ɹ/-/l/-/w/). They instead pay primary attention to variation in F2 frequency and phonemic duration as an interlanguage strategy, resulting in non-nativelike perceptual representations of both /ɹ/ and /l/ (Iverson et al., 2003). Thus, Japanese learners' ability to accurately perceive /ɹ/ depends on the degree to which they are sensitive to F3 variation.

In terms of production, General American English /ɹ/ is realized through simultaneous constrictions of the lips, the palatal and pharyngeal regions of the vocal tract. For the latter type of constrictions, native speakers of English (NEs) use both retroflexed /ɹ/ (i.e., raising the tongue tip and lowering the tongue dorsum) and bunched /ɹ/ (i.e., lowering the tongue tip and raising the tongue dorsum) (Dellatre & Freeman, 1968). Importantly, regardless of either retroflexed or bunched /ɹ/, such articulatory gestures interact to create a front cavity resonance behind the sublingual space, which is strongly related to the low F3 frequency (i.e., F3 < 2400 Hz for /ɹ/) (Espy-Wilson et al., 2000).[2] In an analysis on speech sound production in Japanese and English, Lotto, Sato, and Diehl (2004) found that NJs tend to produce English /ɹ/ via tongue retraction (lower F2 values) without resorting to articulatory parameters (i.e., lower F3 for labial, palatal, and pharyngeal constrictions).

Taken together, these previous studies provide a suggestive pattern for Japanese speakers' interlanguage acquisition of English /ɹ/ both in perception and production. First, Japanese learners perceive and produce English /ɹ/ using the "NJ default strategy" of F2 (1700-2100 Hz → 900-1500 Hz) and duration (5-20ms →50-100ms) dependency. Second, they need to pay an increasing amount of attention to new, unfamiliar parameters such as F3 variance (2400-3000 Hz → 1600-1900 Hz), resulting in change in orolingual articulation that includes narrowed labial constriction (for word-initial tokens) and an adequate sublingual cavity (for F3 resonance). For work on the modeling of English /ɹ/ acquisition in naturalistic settings, see Saito, 2013; Saito and Brajot, 2013.

Current Study

The current study assessed the participants' perception abilities by asking them to listen to distinguish word-initial /ɹ/ (i.e., low F3) from /l/ (i.e., high F3) of English minimal pairs using an aural-mode, two-alternative forced choice identification task. The analysis examines how Japanese learners' perception ability is differentially related to global (perceived accuracy and intelligibility) and specific (F3, F2, F1 and duration) domains of their production ability under

[2] Speech constitutes acoustic signals generated by movements of speech gestures; such acoustic signals have traditionally been described via the first three frequency ranges of energy concentration in Hz (F1, F2, F3). These signals can be used as an index of how talkers actually use their articulators to produce sounds (F1 for degree of tongue height, F2 for tongue retraction, F3 for labial, palatal, and pharyngeal constrictions).

different task conditions (word reading vs. picture description). As such, the ultimate goal of the project was to reveal whether, to what degree and how such perception ability could differentially predict various stages of interlanguage pronunciation development: (a) the resetting of the existing articulatory parameters (the rate and degree of tongue retraction for low F2 and long phonemic duration); (b) the acquisition of the new articulatory parameters (labial, palatal and pharyngeal constrictions for low F3); and (c) automatization (a transition from controlled to spontaneous processing).

Participants

Japanese learners of English. In previous perception-production studies (e.g., Bradlow et al., 1997; Goto, 1971; Sheldon & Strange, 1982), participants' proficiency levels were typically restricted to either "inexperienced" or "experienced" L2 learners, limiting the generalizability of their findings to other learners at various stages of L2 speech learning. To this end, the participants were carefully selected, as they were to reflect a wide range of L2 pronunciation proficiency levels that late Japanese learners of English could generally attain. The 45 Japanese (39 females and 6 males) participants were either students at a private language institute in Osaka, Japan or volunteers from neighboring colleges/universities (age: $M = 30.08$ years, $SD = 8.12$ years). All of the participants had received six years of English education in a foreign language context. None of them reported any extensive experience in pronunciation training—a typical learner profile especially in foreign language classrooms in Japan (for the details of the lack of learner awareness about the importance of pronunciation in L2 communication as well as teacher training on pronunciation teaching, see Saito, 2014; Saito & van Poeteren, 2012). As summarized in Table 1, their previous length of residence in an English-speaking environment (the US) was varied ($M = 6.82$ months, $SD = 9.87$ months). According to their English test scores (TOEIC), their proficiency levels widely varied from 400 (beginner levels) to 950 (highly advanced levels).[3]

Table 1.
Participants' previous length of residence in the US

Length of residence in the US	No. of participants
0	23
$0 < x < 1$ year	14
$1 < x < 3$ years	8

Native baselines. For the sake of comparison, 10 native speakers of Canadian English were recruited to complete the same perception and production tests (age: $M = 25.30$ years, $SD = 6.81$ years). All were graduate or undergraduate students at an English-speaking university in Montreal at the time of the project.

Native raters. Five listeners (three males, two females) were recruited from undergraduate linguistics and psychology courses at a Canadian university in Vancouver. All participants were native speakers of Canadian English: All of the raters were from Vancouver (age: $M = 25.2$ years, $SD = 3.41$ years). They reported little familiarity and contact with

[3] A total of 35 participants' proficiency scores were available. The other participants ($n = 10$) had never taken TOEIC prior to the current project.

Japanese-accented English (M = 1.5, range = 1-2: 1 = *Not at all,* 6 = *Very much*). All passed a pure-tone screening at octave frequencies between 250 and 4000 Hz.[4]

Outcome Measures

All testing sessions were carried out individually in a quiet room. To avoid drawing the participants' attention to the target segmental sound (English /ɹ/), especially in the spontaneous production task, both the Japanese learners and native baselines took the perception and production tests in the following order: (a) spontaneous production, (b) controlled production and (c) perception. For the production tests, recordings were made with a Yamaha unidirectional DM-20SL microphone and a Roland R05 Wave recorder (44.1 kH, 16-bits) in a quiet room. Audio files were saved on a laptop for subsequent analysis. All participants were paid for their participation.

Production Data

Spontaneous production test. The current study adopted the timed picture description task used in previous relevant L2 phonology research (e.g., Rau et al., 2009), but with some modifications. In particular, efforts were made to ensure that participants produced /ɹ/ under communicative pressure, but without accessing orthographic representations of the target words. The test was administered as follows:

1. NJs were given 10 seconds to memorize a written list of four key words.
2. Immediately afterwards, the list was taken away. Then, the participants were asked to describe two pictures in a row with no planning time using two of the key words for each picture, one of which was a target word that included /ɹ/ in the word-initial position.
3. After describing the pictures, they moved on to the next four key words for another pair of pictures.

In total, the participants described 10 pictures with target words and six distracter pictures. Since NJs produce word-initial /ɹ/ in a significantly different manner according to the degree of backness of the subsequent vowel (Bradlow et al., 1997), this variable was carefully controlled (see Table 2 for the following vowel conditions of the 10 target tokens). According to Vocabulary Profiler (Cobb, 2010), all target lexical items fell within the first 2000 word families except for "route" (included in the academic word list). All materials were adapted from Saito (2015).

[4] It is noteworthy that the native baseline data was produced by native speakers of North-western Canadian English (i.e., residents in Vancouver), and rated by those of "North-eastern" English (i.e., residents in Montreal). Canadian English is spoken in a vast area (from British Columbia to Quebec), and there is some phonological phenomenon specific to this generic variety relative to other dialects of North American English, such as Canadian Raising (i.e., the diphthongs /aɪ, aʊ/ are raised to /ʌɪ, ʌʊ/ before voiceless consonants) (Chambers, 1973). According to Brinton and Fee's (2001) review, there is "a remarkable homogeneity in speech" across the country. Therefore, I used the term, Canadian English, to refer to the dialect of English the participants in the current study used at the time of the project.

Table 2.
10 tokens in the spontaneous production test in relation to following vowel conditions

	Front vowels	Central vowels	Back vowels
High vowels	Read /ɹid/, ring /ɹɪŋ/		Roof /ɹuf/, route /ɹut/
Mid vowels	Rain /ɹeɪn/, red /ɹɛd/	Rush /ɹʌʃ/	Road /ɹoʊd/, rope /ɹoʊp/
Low vowels		Rice /ɹaɪs/	

Controlled production test. In order to elicit controlled production of /ɹ/, the participants were asked to read a list of 40 words, out of which 15 were target tokens. All target lexical items fell within the first 2000 word families according to Vocabulary Profiler (Cobb, 2010) (i.e., these words are highly frequent) except for "Ryan" and "rink."[5] As shown in Table 3, the following vowels of the 15 target tokens were controlled and considered roughly similar to those of the 10 target tokens in the spontaneous test.

Table 3.
15 tokens in the controlled production test in relation to following vowel condition

	Front vowels	Central vowels	Back vowels
High vowels	Rink /ɹɪŋk/, reef /ɹif/, reach /ɹitʃ/		Rule /ɹul/, room /ɹum/, rude /ɹud/
Mid vowels	Race /ɹeɪs/, rent /ɹɛnt/, rate /ɹeɪt/	Rough /ɹʌf/	Road /ɹoʊd/, wrong /ɹɔŋ/, roll /ɹoʊl/
Low vowels		Ryan /ɹaɪən/, right /ɹaɪt/	

Acoustic analysis. Following the acoustic analysis used for natural speech tokens of word-initial /ɹ/ established by Flege (Flege, Takagi, & Mann, 1995) and extended by Saito (Saito, 2013; Saito & Brajot, 2013), the researcher measured F3, F2 and duration values of all speech tokens using the speech analysis software *Praat* (Boersma & Weenik, 2012). The following procedure was used:

1. The researcher listened to each speech token and determined the beginning of the word by using spectrographic representations and wave forms.
2. With the view of identifying the beginning of word-initial /ɹ/ embedded in continuous speech (spontaneous production) as a reliable cue, the researcher looked for the endpoint of falling F3. Another reliable cue is the onset of the steady state of F1, F2, and F3 (i.e., the F3 of the preceding sounds tends to continue to decline towards the beginning of the word because the F3 of /ɹ/ is relatively low). F3 and F2 values were calculated at this point.

[5] Both monothongs (e.g., *read, rock*) and diphthongs (e.g., *Ryan, right*) were classified into the same category according to following vowel conditions in the current study. Furthermore, "Ryan" is a bisyllabic word although the other items are monosyllabic words. However, no separate analyses were further conducted for monothongs and diphthongs, given that no empirical research has ever examined how Japanese learners' /ɹ/ production varies when preceding monothongs or diphthongs.

3. Finally, phonemic duration was measured by dragging a cursor from the beginning point of the F1 transition to the endpoint of the F1 or F3 transition.

Normalization. Spectral information (i.e., F3, F2, F1 values) significantly varies due to anatomical differences in individual vocal tract length (e.g., ±15cm). Thus, for comparison purposes, raw acoustic values were adjusted via the following normalization procedure on spectral information in human speech, as suggested by Lee, Guion, and Harada (2006). First, a mean F3 value of /æ/ elicited from 10 monosyllabic words in the controlled production test (i.e., *fan, tap, map, bat, mad, cap, bag, lab, cat, dad*) was calculated for each participant ($N = 45$ Japanese learners + 10 NS baseline speakers). One female native English speaker was randomly selected as a reference, and her mean F3 value (i.e., 2897 Hz) was divided by those of the other participants to provide their own *k* factors. Then, all formant values (F3, F2, F1) of /ɹ/ for each participant were multiplied by their respective *k* factor. All acoustic values in Hertz were converted into Bark using the formula described in the *Praat* manual (Boersma & Weenink, 2011). This second adjustment is considered to reduce the nonlinear relationship between the formant frequencies and the corresponding perceived semivowel quality.[6]

$$Bark = 7\ln\left(\frac{Hz}{650} + \sqrt{1 + \left(\frac{Hz}{650}\right)^2}\right)$$

Listener judgements. A total of 1375 tokens (25 /ɹ/ tokens [10 spontaneous tokens + 15 controlled tokens] × 55 participants [45 Japanese learners + 10 native baselines]) were normalized for peak intensity and saved as .wav files. For the spontaneous /ɹ/ tokens, the researcher carefully listened to the speech samples multiple times to avoid any significant distortion of the extracted words.

The listening sessions were conducted individually in a quiet room at the Canadian university. All speech tokens were presented to the five listeners on a laptop computer screen. To reduce their fatigue, they took a 10-minute intermission halfway through (the entire session took approximately 2.5 hours). A 9-point scale descriptor was adapted and modified from Flege, Takagi, and Mann's (1995) 6-point scale. Upon hearing each token, the listeners clicked one of the nine rating criteria: 1 (*very good /ɹ/*) → 2 (*good /ɹ/*) → 3 (*probably /ɹ/*) → 4 (*possibly /ɹ/*) → 5 (*neutral exemplars, neither /ɹ/ nor /l/*) → 6 (*possibly /l/*) → 7 (*probably /l/*) → 8 (*good /l/*) → 9 (*very good /l/*).

The listeners could use a "repeat" button to hear an item up to three times until they felt satisfied about their judgement. They were explicitly asked to focus their judgements on only the quality of /ɹ/; this was because their ratings would have otherwise been affected by the quality of the entire word, likely other pronunciation errors typical of Japanese learners (e.g., /ɹi̠ːŋ / for "ring").

First, they rated five speech tokens (not included in the subsequent listening session) on a 9-point scale as practice. Their 9-point scores were used without any modification for accuracy (to what degree the participants' /ɹ/ production is targetlike), and calculated for intelligibility (1 [*very good /ɹ/*] – 4 [*possibly /ɹ/*] were recorded as intelligible /ɹ/ exemplars; 5-9 [neutral or /l/] as

---

[6] Bark is a conversion formula used to adjust raw acoustic values in Hz to human perception range, because frequency range is not linear (e.g., a change from 400 to 500 Hz is much more perceptible than a change from 2000-2100 Hz). In this sense, all raw acoustic values were converted into Bark so that any change in Bark values (e.g., 5 to 6 Bark, 10 to11 Bark) would have the same impact on human perception.

unintelligible /ɹ/). According to Cronbach alpha analyses, the five raters' inter-rater reliability was relatively high for accuracy (α = .92) and intelligibility (α = .93). By pooling over listeners, mean rating scores were assigned to each /ɹ/ token produced by the Japanese learners.

Perception Tests

Both the Japanese learners and English baselines completed the two-alternative, forced choice identification task which consisted of 50 minimally-paired words. The 70 test items included 50 minimally-paired words beginning with word-initial /ɹ/ or /l/ ("rain" vs. "lane") together with 20 distracter minimally-paired words (e.g., "think" vs. "sink"). The words were spoken by one male native speaker of Canadian English, and were recorded in isolation in a quiet room at an English speaking university in Montreal. The tokens were digitized at a 40-kHz sampling rate and normalized for peak intensity using the *Praat* (Boersma & Weenik, 2010) speech analysis software.

The set of target minimally-paired words (n = 50) contrasted word-initial /ɹ/ and /l/ in three phonetic contexts: 20 singletons with front vowels, 10 singletons with central vowels, and 20 singletons with back vowels (see Table 4). Though several words were infrequent (e.g., lush, loll), the participants were told that all items were real words (without any nonsense ones), and that they would listen to either minimally-paired or "nearly" minimally paired words. They were explicitly asked to focus on the contrasting sounds (e.g., /ɹ/-/l/, /θ/-/s/, /v/-/b/) rather than lexical meaning while identifying the target phoneme (/ɹ/ vs. /l/).

Table 4.
50 tokens in the perception tests in relation to following vowel conditions

	Front vowels	Central vowels	Back vowels
High vowels	"rink /ɹɪŋk/, link /rɪŋk/" "reef /ɹif/, leaf /rif/" "read /ɹid/, lead /lid/" "ring /ɹɪŋ/, ling /lɪŋ/" "reach /ɹitʃ/, leach /litʃ/"		"roof /ɹuf/, Loof /luf/" "rule /ɹul/, lure /lʊər/" "room /ɹum/, loom /lum/" "rude /ɹud/, lude /lud/" "root /ɹut/, loot /lut/"
Mid vowels	"race /ɹeɪs/, lace /leɪs/" "rent /ɹent/, lent /lent/" "rain /ɹeɪn/, lane /leɪn/" "rate /ɹeɪt/, late /leɪt/" "red /ɹɛd/, led /lɛd/"	"rough /ɹʌf/, laugh /lʌf/" "rush /ɹʌʃ/, lush /lʌʃ/"	"road /ɹoʊd/, load /roʊd/" "wrong /ɹɔŋ/, long /lɔŋ/" "roan /ɹoʊn/, loan /loʊn/" "roll /ɹoʊl/, loll /loʊl/" "rope /ɹoʊp/, lope /loʊp/"
Low vowels		"right /ɹaɪt/, light /laɪt/" "ride /ɹaɪd/, lied /laɪd/" "rice /ɹaɪs/, lice /laɪs/"	

Right after the two production tests, the perception test was administered. The participants listened to test items via a laptop computer, and responded to each speech token by marking one of two orthographic choices on a prepared answer sheet.

Results

NJs and NEs' /ɹ/ Performance

As summarized in Table 5, descriptive statistics indicate that NJs' perception and production performance greatly varied, and substantially different from that of the native speaker baseline. Given that multiple dimensions of production were comprehensively analyzed, a set of two-way ANOVAs were conducted to further examine the extent to which the participants' production scores (accuracy, intelligibility, F3, F2, F1, and duration) differed according to the task (spontaneous vs. controlled) and talker (NJs vs. NEs) conditions. The results yielded significant talker effects for (a) accuracy $F(1, 53) = 238.97, p < .001$, (b) intelligibility $F(1, 53) = 252.04, p < .001$, (c) F3, $F(1, 53) = 161.78, p < .001$, (d) F2, $F(1, 53) = 576.65, p < .001$, (e) F1, $F(1, 53) = 292.43, p < .001$, and (f) duration, $F(1, 53) = 607.06, p < .001$. Significant task effects were found only for duration, $F(1, 53) = 18.246, p < .001$. According to Bonferroni multiple comparisons, NEs' /ɹ/ production was significantly longer in the controlled production test (M = 88.48ms) than in the spontaneous production test (M = 65.61) ($p < .001$), though such a task effect was not found for NJs' phonemic duration ($p > .05$).[7]

Table 5.

Descriptive results of spontaneous production, controlled production and perception test scores

	NJs (*n* = 45)				NEs (*n* = 10)			
	M	*SD*	*Range* Max	*Range* Min	*M*	*SD*	*Range* Max	*Range* Min
A. Spontaneous production								
Accuracy[a]	4.21	1.29	6.82	2.31	1.00	*n.a.*	*n.a.*	*n.a.*
Intelligibility[b]	57.9%	30.2	100%	3.8%	100%	*n.a.*	*n.a.*	*n.a.*
F3	14.57	0.59	15.78	13.48	11.52	0.62	12.75	10.89
F2	11.39	0.74	13.13	9.72	9.82	0.72	11.01	9.04
F1	4.49	0.47	5.51	3.41	4.10	0.50	5.10	3.36
Duration	28.93	13.71	53.20	10.00	65.61	12.22	86.10	48.30
B. Controlled production								
Accuracy[a]	4.28	1.54	7.21	2.16	1.00	*n.a.*	*n.a.*	*n.a.*
Intelligibility[b]	57.5%	35.2	98.8%	2.5%	100%	*n.a.*	*n.a.*	*n.a.*
F3	14.25	0.66	16.22	12.85	11.44	0.55	12.20	10.88
F2	11.31	1.02	13.31	8.75	9.52	0.61	10.19	8.57
F1	4.29	0.48	5.41	3.28	3.81	0.49	4.53	3.07
Duration	28.33	14.58	59.95	10.00	88.48	16.58	121.22	69.95
C. Perception								
Correct scores (50 questions)	67.2%	14.2%	46%	96%	100%	*n.a.*	*n.a.*	*n.a.*

Note. [a]Accuracy was based on a 9-point scale (*1 = good /ɹ/, 9 = good /l/*); [b]Intelligibility was based on a dichotomous category (*/ɹ/ or non-/ɹ/*)

[7] Although the task effect was significant for F1, $F(1, 53) = 12.748, p = .001$, this could be due to the fact that the controlled task included more mid and high vowels following /ɹ/ (resulting higher F1 values) than the spontaneous task did (see Tables 2 and 3).

Perception-Production Link

To examine to what degree NJs' perception scores could be correlated with the five dimensions of their production scores (accuracy, intelligibility, F3, F2, F1 and duration), a set of Spearman nonparametric correlation analyses were performed (see Table 6). First, perception scores were strongly related to overall judgements of spontaneous production performance (r = -.405 for accuracy, r = .432 for intelligibility) and controlled production performance (r = -.628 for accuracy, r = .589 for intelligibility). Second, the predictive power of perception scores for the acoustic properties of English /ɹ/ production slightly varied according to the task conditions (spontaneous vs. controlled). Whereas the perception scores were associated with F2 (r = -.379) and duration (r = .301) in the controlled task, they were only associated with duration (r = .589) in the spontaneous task (r = .320). Interestingly, NJs' perception ability was not significantly predictive of their F3 values in any contexts (i.e., the index for their mastery of the new and primary articulatory parameters for producing targetlike English /ɹ/).

Table 6.
Spearman correlations between perception scores and spontaneous and controlled production scores

Production scores	Perception scores	
	r	*p*
A. Spontaneous production		
Accuracy	-.405	.006**
Intelligibility	.432	.003**
F3	-.005	.974
F2	-.188	.216
F1	.064	.676
Duration	.320	.032*
B. Controlled production		
Accuracy	-.628	< .001**
Intelligibility	.589	< .001**
F3	-.189	.213
F2	-.379	.010**
F1	.139	.363
Duration	.301	.044*

* indicates $p < .05$; ** indicates $p < .01$

Acoustic Correlates of English /ɹ/

The final objective of the statistical analyses was to examine the acoustic correlates of NJs' English /ɹ/ production. The relationship between the overall quality (accuracy, intelligibility) and the acoustic properties of the NJs' /ɹ/ production was analyzed via a set of stepwise multiple regression analyses. Accuracy and intelligible scores served as dependent variables, with the four acoustic variables (F3, F2, F1 and duration) used as predictors.

Spontaneous production. The regression model, which included two acoustic variables (F3, duration), accounted for 71.1% of the variance in accuracy and 51% in intelligibility, with no evidence of strong collinearity in the model (VIF = 1.72). Whereas native raters used F3 as a primary cue (66%) and duration as a secondary cue (5%) for the accuracy judgements, the reverse pattern was found for the intelligibility judgements (3% for F3, 48% for duration).

Table 7.
Results of multiple regression analysis using acoustic variables as predictors of spontaneous /ɹ/ production

Predicted variable	Predictor variables	*Adjusted* R^2	R^2 *change*	*F*	*p*
Accuracy	F3	.66	.66	108.52	< .001
	Duration	.71	.05	67.41	< .001
Intelligibility	Duration	.48	.48	51.45	< .001
	F3	.51	.03	29.49	< .001

Note. The variables entered into the regression equation included F3, F2, F1 and duration.

Controlled production. The regression model identified different predictors for native raters' global judgements of NJs' /ɹ/ production according to rated categories (accuracy vs. intelligibility). Whereas much variance in their accuracy judgements was explained primarily by F3 (64%) and secondarily by F2 (3%), their intelligibility judgements were strongly tied with F2 (45%) and duration (7%). The model did not demonstrate strong collinearity (*VIF* = 1.63)

Table 8.
Results of multiple regression analysis using acoustic variables as predictors of controlled /ɹ/ production

Predicted variable	Predictor variables	*Adjusted* R^2	R^2 *change*	*F*	*p*
Accuracy	F3	.64	.64	95.93	< .001
	F2	.67	.03	55.15	< .001
Intelligibility	F2	.45	.45	44.59	< .001
	Duration	.52	.07	30.72	< .001

Note. The variables entered into the regression equation included F3, F2, F1 and duration.

Discussion

Whereas major L2 speech theories (e.g., Best & Tyler, 2007; Flege, 2003) have generally assumed that the perception and production domains are interrelated in the process and product of L2 speech learning, it still remains unclear the extent to which such perception ability can predict multiple dimensions of L2 speech production ability. In the context of Japanese learners' English /ɹ/ performance, the current investigation set out to expound the predictive power of perception ability for four different dimensions of production ability: (a) global quality (native speakers' overall judgements); (b) the redeployment of existing articulatory parameters (the rate and degree of tongue retraction); (c) the acquisition of new articulatory parameters (labial, palatal and pharyngeal constrictions); and (d) automatization (a transition from controlled to spontaneous processing).

First and foremost, the results demonstrated that NJs' /ɹ/ perception ability, measured via a two-alternative forced choice identification task, was correlated with the global judgements of their /ɹ/ production ability. These results concur with the majority of L2 phonology research which has also found a significant perception and production link in other aspects of L2 speech, such as vowels (e.g., Baker & Trofimovich, 2006; Flege et al., 1997) and syllable structures (e.g., Cardoso, 2011; Flege, 1993). Although adult L2 learners are likely to demonstrate different linguistic behaviours according to different task conditions (Major, 2008), the results here suggest that the tie between perception and production could be robust at both the controlled and spontaneous levels.

Importantly, the results of the acoustic analyses further revealed precisely how such perception ability was predictive of NJs' use of different articulatory parameters during their /ɹ/ production. The results found that the participants' perception ability was significantly predictive of their F2 and duration values, but not their F3 values. The stronger influence of perception on F2 and duration rather than on F3 may be associated with the different amount of learning difficulty inherent in the former and latter aspects of English /ɹ/ production development.

In the naturalistic L2 speech learning literature, NJs likely enhance the intelligibility of their English /ɹ/ production by readjusting their existing articulatory parameters (lower F2 and longer duration for the rate and degree of tongue retraction) within the first few years of immersion. In contrast, it requires a great deal of L2 experience (5+ years) for NJs to show any significant change in their use of new articulatory parameters (lower F3 for labial, palatal, and pharyngeal constrictions) (Flege et al., 195; Lotto et al, 2004; Saito, 2013; Saito & Brajot, 2013). As shown in the previous literature (Saito, 2013; Iverson et al., 2003) and the current study, native speakers of English indeed rely on F3 as primary information, and F2 and duration as secondary information to assess the targetlikeness of English /ɹ/ production. Therefore, the results presented here indicate that the perception-production link may be strong particularly in the early (but not necessarily later) phase of L2 speech learning. That is, good perception ability may help L2 learners to achieve more intelligible English /ɹ/ production via the resetting of existing articulatory parameters. However, such perception ability per se may be a necessary but not sufficient condition for adult L2 learners to further improve the accuracy of their already intelligible English /ɹ/ production by acquiring new articulatory parameters.

Interestingly, the asymmetrical relationship between perception, production and experience is highly relevant to studies extensively investigating the nativelikeness of NJs' /ɹ/ acquisition. On the one hand, there is some evidence that many experienced NJs can attain nativelike categorical perception of various key acoustic variations in the English /ɹ/-/l/ contrast such as phonemic duration (Underbakke, Polka, Gottfried, & Strange, 1988) and F3 (MacKain, Best, & Strange, 1981). On the other hand, very few NJs seem to reach nativelike F3 range in their English /ɹ/ production even after extensive length of residence in an L2 speaking environment (Larson-Hall, 2006; Saito, 2013; Saito & Brajot, 2013). Thus, it seems to be reasonable to assume that whereas perception may greatly and quickly change as a function of increased experience, and serve as an anchor for initial production development, the attainment of nativelike production is subject to much variability (Sheldon & Strange, 1982). Rather, the noticing, development and refinement of new articulatory parameters (simultaneous constrictions) could be tied with unique individual difference profiles, such as early age of acquisition (Abrahamsson, 2012) and language aptitude (Granena & Long, 2013).

Whereas the discussion thus far is concerned with naturalistic L2 speech learning, a growing number of scholars have investigated whether and how pronunciation instruction can help L2 learners improve not only "intelligibility" but also "accuracy" dimensions of L2 pronunciation (especially given that the latter construct may otherwise remain unchanged or require a great amount of naturalistic exposure). For example, certain training even without any communicative use of language (e.g., High Phonemic Variability Training) could greatly improve the intelligibility of NJs' English /ɹ/ production (e.g., Bradlow et al., 1997: see also

Thomson, 2012 for vowel training). At the same time, there is ample evidence that when pronunciation-focused teaching is integrated into meaning-oriented instruction contexts (e.g., focus on form via pronunciation-focused recasts: Lee & Lyster, 2016), such balanced syllabus could significantly help NJs master various articulatory dimensions of English /ɹ/ (lower F3 and F2, longer phonemic duration) at both controlled and spontaneous speech levels (Saito, 2015). As shown earlier in the Method section, the participants in the study had not received any extensive pronunciation training—a common practice in the Japanese English education system (Saito, 2014; Saito & van Poeteren, 2012). Thus, one promising direction for future L2 speech research is to explore the extent to which instruction can differentially impact the English /ɹ/ production abilities of NJs with different English /ɹ/ perception abilities, as such future studies would allow us to revisit the findings of the study (with a cross-sectional research design) from a longitudinal perspective.

Limitations

One of the major shortcomings of the current study is that its discussion is exclusively limited to the acquisition of /ɹ/ by NJs in an foreign language setting. More research is needed to test the generalizability of the findings to various kinds of learners (including those with different age profiles) in different learning contexts, as well as for a range of other segmental and suprasegmental features of L2 speech. Second, the two-alternative, forced choice identification task was adopted out of necessity as a way to measure participants' perception performance of /ɹ/. Future studies need to elaborate different perception tasks, such as identification tasks with the synthetic F2 × F3 continua (Iverson et al., 2003) and in noisy conditions (Munro, 1998). In addition, it is necessary to assess the processing speed of accurate identification (Munro & Derwing, 1995) under various lexical contexts (i.e., /ɹ/ in frequent words vs. infrequent words: Flege, Takagi, & Mann, 1996). These tests will reveal a more detailed picture of perception abilities in interlanguage phonology. Last, it needs to be acknowledged that the NJ performance of /ɹ/ was cross-sectionally analyzed in the study. Therefore, future studies need to adopt a longitudinal approach towards illustrating if a group of NJs would similarly or differentially develop their perception and production ability over a long period of time (1-5 years) (cf. Munro & Derwing, 2013).

Conclusion

In the context of the performance of word-initial /ɹ/ by adult NJs with various proficiency levels, the current study aimed to re-examine the relationship between L2 perception and production ability. Overall, their perception scores were correlated with global qualities (accuracy, intelligibility) of production both at controlled and spontaneous speech levels. The acoustic analyses revealed that their perception ability was significantly predictive of their redeployment of existing articulatory parameters (i.e., lower F2 for the rate and degree of tongue retraction), but not of their acquisition of new articulatory parameters (i.e., lower F3 for labial, palatal, and pharyngeal constrictions). These results in turn highlight two crucial implications in regards to the perception-production link in interlanguage phonology. First, the perception-production link may be relatively strong at the early, initial phase of L2 speech learning, as good perception ability may induce L2 learners to improve the intelligibility of their L2 controlled and spontaneous production alike via the re-adjustment of existing articulatory cues (e.g., enhancing rate and degree of tongue retraction for producing English /ɹ/). Second, perception may be weakly related to production in the later phase of L2 speech learning at best. This is arguably because the mastery of nativelike pronunciation via the acquisition of new articulatory cues (e.g., creating three constrictions in the vocal tract for producing English /ɹ/) may entail not only good perception ability, but other individual difference factors as well, such as age of acquisition, language learning aptitude or/and pronunciation instruction experience.

References

Abrahamsson, N. (2012). Age of onset and nativelike L2 ultimate attainment of morphosyntactic and phonetic intuition. *Studies in Second Language Acquisition, 34*, 187-214.

Abrahamsson, N. & Hyltenstam, K. (2009). Age of acquisition and nativelikeness in a second language – listener perception *vs.* linguistic scrutiny. *Language Learning, 59*, 249–306.

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*, 231-250.

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception. In O. Bohn, & M. Munro (Eds.), *Language experience in second language speech learning: In honour of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins.

Boersma, P., & Weenik, D. (2012). *Praat: Doing phonetics by computer*. Retrieved from http://www.praat.org (Last viewed September, 2012).

Bohn, O.-S., & Flege, J. (1997). Perception and production of a new vowel category by second-language learners. In A. James & J. Leather (Eds.), *Second-language speech: Structure and process* (pp. 53-74). Berlin: Mouton de Gruyter.

Brinton, L., & Fee, M, (2000). Canadian English. In J. Algeo (Ed.), *The Cambridge History of English* (pp. 422-440). Cambridge: Cambridge University Press.

Bradlow, A., Pisoni, D., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /ɹ/ and /l/. *Journal of the Acoustical Society of America, 101*, 2299–2310.

Bundgaard-Nielsen, R., Best, C., Kroos, C., & Tyler, M. (2012). Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics, 33,* 643-664.

Cardoso, W. (2011). The development of coda position in second language phonology: A variationist perspective. *Second Language Research*, *27*(3).

Cobb, T. (2010). *The compleat lexical tutor*. Retrieved from http://www.lextutor.ca/vp/

Delattre, P., & Freeman, D. (1968). A dialect study of American r's by x-ray motion picture. *Linguistics*, *44,* 29–68.

Derwing, T. M., & Rossiter, M. J. (2003). The Effects of Pronunciation Instruction on the Accuracy, Fluency, and Complexity of L2 Accented Speech. *Applied Language Learning*, *13*(1), 1-17.

Espy-Wilson, C. (1992). Acoustic measures for linguistic features distinguishing the semivowels /wjrl/ in American English. *Journal of Acoustical Society of America, 92*, 736-751.

Espy-Wilson, C. Boyce, S., Jackson, M., Narayanan, S., & Alwan, A., (2000). Acoustic modeling of American English /ɹ/. *Journal of the Acoustical Society of America*, *108*, 343-356.

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

Flege, J. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319-355). Berlin: Mouton de Gruyter.

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

Flege, J., MacKay, I., & Meador, D. (1999). Native Italian speakers' production and perception of English vowels. *Journal of the Acoustical Society of America*, *106*, 2973-2987.

Flege, J., Takagi, N., & Mann, V. (1995). Japanese adults learn to produce English /ɹ/ and /l/ accurately. *Language and Speech, 38*, 25-55.

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *Journal of Acoustical Society of America, 99,* 1161-1173.

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.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds L and R. *Neuropsychologia*, *9,* 317–323.

Hattori, K., & Iverson, P. (2009). English /ɹ/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America, 125*, 469-479.

Hardison, D. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics, 24,* 495-522.

Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*, 847-857.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics, 26*, 227-247.

Larson-Hall, J. (2006). What does more time buy you? Another look at the effects of long-term residence on production accuracy of English /ɹ/ and /l/ by Japanese speakers. *Language and Speech, 49*, 521-548.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*, 345-366.

Lee, A. H., & Lyster, R. (2016). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition*, *38*, 35-64.

Liberman, A. & Whalen, D. (2000). On the relation of speech to language. *Trends in Cognitive Sciences, 4*, 187-196.

Lively, S., Logan J., & Pisoni, D. (1993). Training Japanese listeners to identify English /ɹ/ and /l/. II: The role of phonetic environments and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*, 1242–1255.

Lotto, A. J., Sato, M., & Diehl, R. L. (2004). Mapping the task for the second language learner: The case of Japanese acquisition of /ɹ/ and /l/. In J. Slifka, S. Manuel, & M. Matthies (Eds.), *From Sound to Sense: 50+ Years of Discoveries in Speech Communication* (pp. C381-C386). Electronic conference proceedings.

MacKain, K.S., Best, C., & Strange, W. (1981) Categorical perception of English /ɹ/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2*, 369-390.

Major, R. (2008). Transfer in second language phonology: A review. In J. Hansen Edwards & M. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 63-94). Amsterdam: John Benjamins.

McAllister, R., Flege, J., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics, 30*, 229-258.

Munro, M. J. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition*, 20, 139-154.

Munro, M., & Derwing, T. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*, 289–306.

Munro, M., & Derwing, T. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System, 34*, 520–531.

Munro, M. & Derwing, T. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning, 58,* 479-502.

Munro, M. J., & Derwing, T. M. (2013). English L2 vowel acqusition over seven years. In. J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*. Ames, IA: Iowa State University

Neufeld, G. (1988). Phonology asymmetry in second-language learning and performance. *Language Learning, 38*, 531-559.

Piske, T., Flege, J., MacKay, & Meador, D. (2011). *I*nvestigating native and non-native vowels produced in conversational speech. In M. Wrembel, M. Kul & Dziubalska-Kołaczyk, K. (Eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010* (pp. 195-205). Switzerland: Peter Lang.

Rau, D., Chang, A., & Tarone, E. (2009). Think or sink: Chinese learners' acquisition of the voiceless interdental fricative. *Language Learning*, *59*, 581-621.

Saito, K. (2013). Age effects on late bilingualism: The production development of /ɹ/ by high-proficiency Japanese learners of English. *Journal of Memory and Language, 69*, 546-562.

Saito, K. (2014). Experienced teachers' perspectives on priorities for improved intelligible pronunciation: The case of Japanese learners of English. *International Journal of Applied Linguistics, 24*, 250-277.

Saito, K. (2015). Communicative focus on L2 phonetic form: Teaching Japanese learners to perceive and produce English /ɹ/ without explicit instruction. *Applied Psycholinguistics, 36*, 377-409.

Saito, K., & Brajot, F. (2013). Scrutinizing the role of length of residence and age of acquisition in the interlanguage pronunciation development of English /ɹ/ by late Japanese bilinguals. *Bilingualism: Language and Cognition, 16*, 847-863.

Saito, K., & Wu, X. (2014). Communicative focus on form and L2 suprasegmental learning: Teaching Cantonese learners to perceive Mandarin tones. *Studies in Second Language Acquisition, 36*, 647-680.

Saito, K., & van Poeteren, K. (2012). Pronunciation-specific adjustment strategies for intelligibility in L2 teacher talk: Results and implications of a questionnaire study. *Language Awareness, 21*, 369-385.

Sheldon, A., & Strange, W. (1982). The acquisition of /ɹ/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3*, 243–261.

Thomson, R. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning, 62*, 1231–1258.

Trofimovich, P., Lightbown, P. M., Halter, R., & Song, H. (2009). Comprehension-based practice: The development of L2 pronunciation in a listening and reading program. *Studies in Second Language Acquisition, 31*, 609-639.

Underbakke, M., Polka, L., Gottfried, T. L., & Strange, W. (1988). Trading relations in the perception of /ɹ/ and /l/ by Japanese learners of English. *Journal of the Acoustical Society of America, 84*, 90–100.

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E., Kawakatsu, M., Tohkura, Y., & Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *Neuroimage, 46*, 226-240.