Laboratory Phonology
Journal of the Association for Laboratory Phonology

Yazawa, K., Konishi, T., Whang, J., Escudero, P., & Kondo, M. (2023). Spectral and temporal implementation of Japanese speakers' English vowel categories: A corpus-based study. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 14(1), pp. 1–33. DOI: https://doi.org/10.16995/labphon.6427

Open Library of Humanities

Spectral and temporal implementation of Japanese speakers' English vowel categories: A corpus-based study

Kakeru Yazawa*, Faculty of Humanities and Social Sciences, University of Tsukuba, Ibaraki, Japan, yazawa.kakeru.gb@u.tsukuba.ac.jp

Takayuki Konishi, Global Education Center, Waseda University, Tokyo, Japan, tkonishi@aoni.waseda.jp

James Whang, Department of Linguistics, Seoul National University, Seoul, South Korea, jamesw@snu.ac.kr

Paola Escudero, The MARCS Institute for Brain, Behaviour & Development, Western Sydney University, Sydney, NSW, Australia; Australian Research Council Centre of Excellence for the Dynamics of Language, The Australian National University, Canberra, ACT, Australia, paola.escudero@westernsydney.edu.au

Mariko Kondo, School of International Liberal Studies, Waseda University, Tokyo, Japan; Graduate School of International Culture and Communication Studies, Waseda University, Tokyo, Japan, mkondo@waseda.jp

*****Corresponding author.

This study investigates the predictions of second language (L2) speech acquisition models — SLM(-r), PAM(-L2), and L2LP — on how native (L1) Japanese speakers implement the spectral and temporal aspects of L2 American English vowel categories. Data were obtained from 102 L1 Japanese speakers in the J-AESOP corpus, which also includes nativelikeness judgments by trained phoneticians. Spectrally, speakers judged to be non-nativelike showed a strong influence from L1 categories, except L2 /ʌ/ which could be deflected away from L1 /a/ according to SLM(-r) and L2 /ɑ:/ which seemed orthographically assimilated to L1 /o/ according to PAM(-L2). More nativelike speakers showed vowel spectra similar to those of native English speakers across all vowels, in accordance with L2LP. Temporally, although speakers tended to equate the phonetic length of English vowels with Japanese phonemic length distinctions, segment-level L1-L2 category similarity was not a significant predictor of the speakers' nativelikeness. Instead, the implementation of prosodic-level factors such as stress and phrase-final lengthening were better predictors. The results highlight the importance of suprasegmental factors in successful category learning and also reveal a weakness in current models of L2 speech acquisition, which focus primarily on the segmental level. Theoretical and pedagogical implications are discussed.

1 Introduction

Second language (L2) speech production is influenced by the phonology of the speaker's already-acquired first language (L1), which often makes adult L2 learners' speech different from that of native speakers. Currently dominant models of L2 speech acquisition (Chen & Chang, 2022) — namely, the Speech Learning Model and its revision (SLM(-r); Flege, 1995; Flege & Bohn, 2021), the Perceptual Assimilation Model and its extension to L2 learning (PAM(-L2); Best, 1995; Best & Tyler, 2007), and the Second Language Linguistic Perception model (L2LP; Escudero, 2005; Escudero & Yazawa, in press; van Leussen & Escudero, 2015) — generally agree that L2 learners may nonetheless approach nativelike productions, hypothetically through the mechanism of phonetic and/or phonological category formation and modification that remains available throughout their lifetime. However, the models do not fully agree on the likelihood and magnitude of such changes. SLM(-r) claims that the perceived phonetic dissimilarity between L1 and L2 sounds predicts category formation and nativelikeness; 'dissimilar' L2 sounds that are perceptually distinct from L1 sounds are likely to undergo new category formation and can consequently become nativelike, while L2 sounds that are 'similar' to L1 sounds are subject to equivalence classification and likely remain non-nativelike. PAM(-L2) aligns with this prediction, although the model differs from SLM(-r) in focusing on sound contrasts rather than individual sounds and in considering both phonetic and phonological levels rather than the allophonic level. In contrast, L2LP claims that nativelike performance can be achieved for both 'dissimilar' and 'similar' L2 sounds and contrasts; all L2 categories can be modified extensively regardless of L1-L2 similarity because they are independent 'copies' of L1 categories. The present study aims to investigate whether category assimilation, formation, and modification as proposed by the three models can adequately explain L1 Japanese speakers' production of L2 American English (AmE)[1] monophthongs /iː ɪ ɛ æː ʌ ɑː uː ʊ/[2] in relation to their L1 vowel productions.

Japanese and AmE vowels are of interest because Japanese has fewer vowel qualities than AmE, with some AmE vowels being spectrally 'similar' to Japanese ones while others being noticeably 'dissimilar.' The L1-L2 (dis)similarities also extend to the temporal domain, as Japanese has phonological vowel length while AmE vowel duration varies phonetically. Thus, the present study serves as a test of whether perceptual similarities between L1 and L2 sound categories or categorical contrasts would predict the production of L2 vowels (according to SLM(-r) and PAM(-L2)) or not (according to L2LP) in both spectral and temporal domains. Another merit in

[1] AmE is chosen as the target variety of English because it is widely used in the formal English language education in Japan and therefore is the most familiar to the learners (Sugimoto & Uchida, 2020).

[2] AmE /eɪ/ and /oʊ/, though also considered monophthongs phonologically, are not investigated in the present study. This is because these phonetically diphthongal vowels are categorized as a sequence of two different vowels in Japanese, namely /ei/ and /ou/ (Strange et al., 1998), which can undergo phonological lengthening under certain conditions that would require a separate analysis. See Yazawa (2022) instead for an analysis of these vowels.

investigating the two languages is that they differ substantially in their prosodic organizations. Current models of L2 speech acquisition are primarily concerned with the segmental level but have little to say about prosodic aspects, although both levels are known to affect L2 production (Anderson-Hsieh, Johnson, & Koehler, 1992; Munro & Derwing, 1995). While several studies have examined Japanese speakers' production of AmE vowels (Fox & McGory, 2007; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005; Oh et al., 2011), their scope was somewhat limited as they examined a relatively small number of speakers with similar L2 proficiency levels or short-term longitudinal changes in such speakers, with little reference to their implementation of relevant L1 categories. The present study, therefore, adds to the limited literature by examining 102 Japanese speakers with varying L2 English proficiency levels, using a Japanese-English bilingual speech corpus called J-AESOP (Kondo, Tsubaki, & Sagisaka, 2015).

The remainder of this introductory section is organized as follows. First, the vowel systems of Japanese and AmE are briefly described (Section 1.1). Second, previous studies on Japanese speakers' perception and production of AmE vowels are presented in relation to the theoretical predictions of the three L2 models (Section 1.2). Third, the relevance of prosodic factors to L2 vowel production is discussed (Section 1.3). Finally, several hypotheses regarding Japanese speakers' implementation of L2 AmE vowel categories are proposed (Section 1.4).

1.1. Vowel systems of Japanese and AmE

The vowel systems of Japanese and AmE differ considerably in their use of quality and quantity (Nishi, Strange, Akahane-Yamada, Kubo, & Trent-Brown, 2008). The Japanese system consists of five distinct qualities /i e a o u/, which form five short (1-mora) and long (2-mora) pairs (Keating & Huffman, 1984). Vowel length is phonemic (e.g., /susi/ 'sushi' vs. /suusi/ 'numeral'). Short vowels contrast in height and backness: /i/ is high front, /e/ is mid front, /a/ is low central, and /o/ is mid back. High back /u/ has traditionally been described as unrounded [ɯ], but a recent ultrasound study suggested that it is in fact closer to central rounded [ʉ] or possibly front rounded [ʏ] (Nogita, Yamane, & Bird, 2013). The vowels thus contrast in lip rounding according to phonological backness: The back vowels are rounded (/o u/) while the non-back ones are unrounded (/i e a/). Long vowels are spectrally very similar to their short counterparts but are approximately two to three times longer in duration (Hirata, 2004; Yazawa & Kondo, 2019). The present study transcribes Japanese long vowels with double letters (/ii ee aa oo uu/) because they can phonologically be considered a sequence of identical vowels, which are phonetically realized as [iː eː aː oː ʉ(ʏ)ː].

AmE has a much more dense vowel system than Japanese, with 10 to 11 monophthongs /iː ɪ eɪ ɛ æː ʌ ɑː (ɔː) oʊ uː ʊ/, three diphthongs /aɪ aʊ ɔɪ/, one rhotic vowel /ɝ/, and the schwa /ə/ (Hillenbrand, Getty, Clark, & Wheeler, 1995; Labov, Ash, & Boberg, 2006). The monophthongs, which are the focus of the present study, contrast in height, backness, and roundedness: /ɔː oʊ uː ʊ/ are rounded while the others are unrounded. The monophthongs can also be classified

as either tense /iː eɪ ɑː ɔː oʊ uː/ or lax /ɪ ɛ æː ʌ ʊ/. The present study does not investigate /eɪ/ and /oʊ/ for the reason stated in Note 2. The present study also treats back /ɑː/ and /ɔː/ as a single phoneme /ɑː/ because they are neutralized as such in many dialects of AmE. The high back vowels /uː/ and /ʊ/ are both undergoing fronting across AmE dialects as well. While vowel length is not phonemically contrastive in AmE, there are systematic differences between the intrinsic duration of peripheral vowels /iː æː ɑː uː/ and their centralized counterparts /ɪ ɛ ʌ ʊ/ (Peterson & Lehiste, 1960; Umeda, 1975), but not as prominent as the Japanese long and short vowel contrast. Phonetically long vowels in AmE are of roughly equal duration to Japanese long vowels, whereas phonetically short AmE vowels are not as short as Japanese short vowels (Nishi et al., 2008). Vowel duration serves as a secondary perceptual cue for vowel identity for /æː/-/ɛ/ and /ɑː/-/ʌ/ but not for /iː/-/ɪ/ and /uː/-/ʊ/ (Hillenbrand, Clark, & Houde, 2000). Although somewhat unconventional, the present study transcribes the phonetically long AmE vowels with the 'long' symbol ("ː") to emphasize their length differences from the phonetically short vowels.

Figure 1 shows the average first formant (F1) and second formant (F2) values of the Japanese and AmE vowels that are the focus of the present study. The data were obtained from Nishi et al. (2008), in which four adult male Japanese speakers produced the Japanese vowels in /hVba/ syllables, recorded in isolation ("citation" condition) or within a carrier sentence ("sentence"

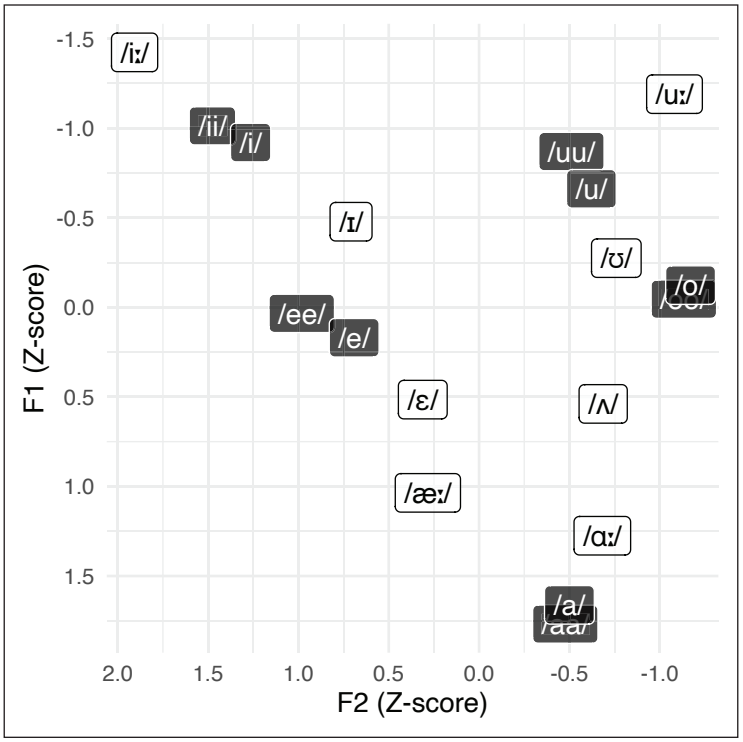

Figure 1: Z-normalized mean F1 and F2 of Japanese (black box) and AmE (white box) vowels of interest, adapted from Nishi et al. (2008).

condition). The study also reports AmE data from a previous study (Strange et al., 1998), in which four male AmE speakers produced the AmE vowels in /hVb(ə)/ syllables in both citation and sentence conditions. The figure shows a Z-normalized version of the production data in the sentence condition, which is useful for comparing the relative spectral (dis)similarities between the Japanese and AmE vowels of interest.

1.2. Previous studies on Japanese speakers' perception and production of AmE vowels

According to SLM(-r), PAM(-L2), and L2LP, perception precedes or co-evolves with production in L2 speech acquisition. Strange et al. (1998) investigated how Japanese speakers categorize AmE vowels into Japanese ones during perception, which would help predict how these speakers produce the L2 AmE vowels. In the study, 24 native Japanese speakers listened to four male native AmE speakers' production of AmE monophthongs in /hVb/ syllables embedded in a short carrier sentence. The Japanese speakers then selected the Japanese vowel category to which each AmE vowel was most similar, also rating its category goodness on a seven-point scale. **Table 1** summarizes the result for the eight AmE monophthongs /iː ɪ ɛ æː ʌ ɑː uː ʊ/ that are of interest to the present study. The results show that phonetically long and short vowels in AmE were systematically assimilated to phonologically long and short Japanese vowel categories, respectively, that match in spectral quality. However, it should be noted that not all L2 sounds were perceived as equally good exemplars of the Japanese categories. In general, tense AmE vowels were rated as better exemplars of Japanese vowel categories than lax vowels. In particular, lax /æː/ was the least consistently assimilated and received the lowest goodness ratings.

AmE		Japanese	Goodness	AmE		Japanese	Goodness
/iː/	→	/ii/	6	/ɪ/	→	/i/	4
/æː/	→	/aa/	2	/ɛ/	→	/e/	4
/ɑː/	→	/aa/	5	/ʌ/	→	/a/	4
/uː/	→	/uu/	5	/ʊ/	→	/u/	3

Table 1: Perceptual categorization of AmE vowels into Japanese vowels with modal responses and median goodness ratings (1 = worst, 7 = best), adapted from Strange et al. (1998).

SLM(-r) predicts that perceptual similarities between L1 and L2 sounds, as represented by the goodness ratings in the above study, affect the likelihood of new L2 category formation. For example, the above data suggest that AmE tense /iː/ and lax /ɪ/ are categorized as Japanese long /ii/ and short /i/, respectively, and thus Japanese learners of English would produce the two AmE vowels with peripheral, overlapping spectral qualities and a large durational difference.

However, since AmE /ɪ/ is not a perfect match to Japanese /i/ (i.e., 'dissimilar'), the learner may notice the phonetic difference between the L1 and L2 sounds and establish a new category for the L2 sound. The property of the newly formed L2 category can match the native norm (i.e., more centralized quality and less short duration than Japanese /i/), unless it is 'deflected' away from an existing L1 category since L1 and L2 categories are considered to exist in a common phonetic space. For example, the new L2 /ɪ/ category could be deflected away from the L1 /e/ or /ee/ category to maintain sufficient phonetic contrast in the common space, whereby the quality of the L2 vowel may not be as centralized as that of native AmE production. On the other hand, AmE /iː/ is less subject to category formation due to its proximity to Japanese /ii/ (i.e., 'similar'). Such an L2 sound is expected to remain non-nativelike because an L1-L2 composite or 'diaphone' category will develop for it. AmE /æː/ is the strongest candidate for category formation among all L2 vowels, given its lowest goodness rating according to the above study.

PAM(-L2) makes somewhat similar predictions to SLM(-r). According to the model, Japanese speakers' categorization of AmE /iː/ and /ɪ/ would fall into a Two-Category assimilation pattern, in which each L2 category is perceived as equivalent to a different L1 category (Japanese /ii/ and /i/, respectively) in a common L1-L2 space. Since learners have little difficulty in discriminating minimally contrasting words for these sounds, no further perceptual learning is likely to occur. This suggests that Japanese speakers are likely to retain L1-like, duration-based distinction of AmE /iː/-/ɪ/. However, the model also proposes an alternative possibility that an L2 category is phonologically assimilated to but phonetically dissimilable from an L1 category. For example, Japanese speakers may equate AmE /ɪ/, which is again not a perfect exemplar of Japanese /i/ phonetically, with the L1 category at the phonological (functional or lexical) level. In such case, a new category can be formed for the L2 sound, possibly leading to a more spectra-based and less duration-based distinction from AmE /iː/ that should remain assimilated to Japanese /ii/. Another common type of perceptual assimilation is the Category-Goodness pattern, in which two L2 categories (e.g., AmE /æː/ and /ɑː/) are perceived as equivalent to a single L1 category (e.g., Japanese /aa/), but one is perceived as being more 'deviant' than the other. The model predicts that a new phonetic and phonological category is reasonably likely to be formed for the 'deviant' L2 sound, whereas category formation is unlikely for the 'better-fitting' one.

Unlike SLM(-r) and PAM(-L2), L2LP considers that L1-L2 perceptual similarity does not play a major role in the implementation of L2 categories. According to the model, L2 learners start with a 'copy' of their acquired L1 categories (*Full Copying* hypothesis), through which L2 sounds are perceived. The model thus does not assume a common L1-L2 space like the other two models and instead presupposes an independent phonological grammar for each of the two languages. The duplicated categories in the L2 grammar are then gradually modified based on the linguistic input until they become optimal for the L2 (optimal perception hypothesis). While different types of learning scenarios are proposed depending on the relationship between L1 and

L2 categories (SIMILAR, SUBSET, and NEW scenarios, in the order of least to most difficult), it is crucial that the model predicts the attainment of optimal L2 implementation for all scenarios. Therefore, AmE /i:/ and /ɪ/, which are initially copies of /ii/ and /i/ in L1 Japanese speakers' L2 grammar, will equally approach nativelike implementation regardless of how similar they are to the L1 sounds, leading to a primarily spectral distinction with reduced reliance on duration (Yazawa, Whang, Kondo, & Escudero, 2020). By extension, there is nothing particular about the perceptual dissimilarity of L2 AmE /æ:/ from L1 Japanese categories. Given that duration serves as a secondary cue for AmE /æ:/-/ɛ/ and /ɑ:/-/ʌ/ in optimal L2 perception, the learners' reliance on duration would likely remain for these pairs.

Previous studies on Japanese speakers' production of AmE vowels provide mixed evidence for the three models' predictions. Fox and McGory (2007) examined the production of 10 AmE monophthongs /i: ɪ eɪ ɛ æ: ʌ ɑ: oʊ u: ʊ/ by 20 native Japanese speakers living in Alabama and Ohio. The study found that the Japanese speakers produced AmE tense /i: eɪ ɑ: u:/ with a longer duration than their lax counterparts /ɪ ɛ ʌ ʊ/, which is consistent with the quantitative categorization pattern in **Table 1**. Some vowels such as /u: ʊ/ and /ɑ: ʌ/ were spectrally overlapping, which is also consistent with the qualitative categorization pattern in the table. Other studies examined how Japanese speakers' production of AmE vowels would change as a function of L2 training and experience. Lambacher et al. (2005) tested whether a six-week auditory training would affect 20 Japanese speakers' production of AmE /æ: ʌ ɑ: ɔ: ɝ/. Following SLM and PAM, the authors predicted that 'dissimilar' /æ:/ would pose less of a problem for Japanese speakers to acquire than /ɑ: ʌ/ that are both qualitatively 'similar' to Japanese /a/. Consistent with this prediction, the speakers produced /æ:/ with less spectral overlap with adjacent vowels and closer formant values to native AmE productions post-training compared to pre-training, whereas their post-training productions of /ɑ: ʌ/ showed a considerable degree of spectral overlap. Yet, category formation may not always predict how L2 acquisition proceeds. Oh et al. (2011) conducted a one-year longitudinal study on the production of AmE tense /i: eɪ ɑ: u:/ and lax /ɪ ɛ ʌ ʊ/ by native Japanese adults (n = 16) and children (n = 16) living in Texas. They hypothesized that the lax vowels, which are generally less similar to Japanese vowels (i.e., 'dissimilar') than the tense vowels (i.e., 'similar'), would show more improvement based on SLM. Contrary to the hypothesis, the study found changes in children's production of both tense /i: ɑ:/ and lax /ɪ ɛ/ in a year's time. In addition, children learned to produce all eight AmE vowels with no significant spectral differences from the productions of age-matched native AmE-speaking children. These findings seem to align better with L2LP, which predicts that learners can achieve optimal implementation of L2 categories regardless of L1-L2 similarities. Also, adults showed no improvement in the study, which suggests that one year is not sufficiently long from the perspective of adult L2 acquisition. A longer longitudinal study, however, would be more difficult and costly to conduct, and thus the present study adopts a corpus-based, cross-sectional approach as an alternative. The

use of J-AESOP also allows an investigation of L1 categories, which none of the previous studies above examined explicitly.

1.3. Prosodic factors in L2 vowel production

Prosodic factors can also influence the implementation of L2 vowel categories, especially in the temporal domain.[3] Although current models of L2 speech acquisition do not pay much attention to prosody, it is very relevant to the present study given the substantial differences in prosodic organizations between Japanese and AmE. Japanese is a mora-timed language, in which the duration of each mora (usually /(C)V/) tends to be consistent (Port, Dalby, & O'Dell, 1987). Japanese also has lexical pitch accent, in which pitch variation within a word alters its meaning (e.g., /ha'si/ 'chopstick(s)', /hasi'/ 'bridge', /hasi/ 'edge') (Beckman & Pierrehumbert, 1986).[4] The presence or absence of lexical pitch accent does not affect phonological vowel quantity (Warner & Arai, 2001). In contrast, English is a stress-timed language (Grabe & Low, 2002; Ramus, Nespor, & Mehler, 1999), in which stressed syllables occur at quasi-regular intervals. Stressed vowels are produced with longer duration, increased intensity, higher pitch, and more peripheral vowel qualities than unstressed counterparts (Fry, 1955; Gay, 1978; Kochanski, Grabe, Coleman, & Rosner, 2005). While unstressed vowels in English are often described as a schwa (i.e., reduced vowel), unstressed syllables can also have unreduced vowels as nuclei (Fear, Cutler, & Butterfield, 1995). Thus, the duration of a vowel is subject to whether it bears stress or not (e.g., /ɪ/ in /ˈɪn.kriːs/ '(an) increase' vs. /ɪnˈkriːs/ '(to) increase'). Such cross-linguistic differences in prosodic organization are known to affect Japanese speakers' production of L2 English vowels. For example, Kondo (2009) examined the production of stressed and unstressed vowels in English by Japanese learners of English and native English speakers. The study found that, while stressed vowels were generally longer than unstressed vowels in both groups, the native group reduced unstressed vowel durations significantly more than the learner group. This indicates that the Japanese speakers' shortening of unstressed vowels was insufficient, presumably due to the transfer of Japanese mora timing to English stress rhythm. However, Japanese speakers seem to be able to eventually acquire nativelike durational control of unstressed vowels, according to Lee, Guion, and Harada (2006) who investigated early and late Japanese- and Korean-English bilinguals. The authors proposed that Japanese speakers' increased sensitivity to durational cues to differentiate phonologically short and long vowels may have helped the acquisition, unlike Korean learners of English whose L1 does not have phonemic vowel length. Yet, as the study examined the duration ratio of unstressed to stressed vowels, it cannot be judged

[3] Although the present study focuses on prosodic influences on vowel duration, vowel spectra can also be affected. See Yazawa, Ozaki, Short, Kondo, and Sagisaka (2015) for an analysis of lexical stress and vowel spectra in Japanese speakers' English using J-AESOP.

[4] The patterns of pitch accent placement vary across dialects of Japanese (Jun & Kubozono, 2020); this example is based on Tokyo Japanese.

whether the observed nativelike performance resulted solely from proper shortening of unstressed vowels or was also supplemented by stronger lengthening of stressed ones.

Vowel duration is subject to post-lexical prosody as well. One such example is phrase-final lengthening, which is usually defined as the lengthening of a rhyme (nucleus and coda) occurring before the boundary between prosodic constituents, roughly reflecting syntactic boundary strength (Ueyama, 1996). Phrase-final lengthening is considered a language-universal phenomenon and has been reported in various languages including Japanese (Beckman & Pierrehumbert, 1986) and English (Turk & Shattuck-Hufnagel, 2007; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). Nevertheless, Japanese speakers may manifest phrase-final lengthening in L2 English somewhat differently from native English speakers. Ueyama (1996) examined the production of English phrase-final vowels by beginning and advanced Japanese learners of English and native English speakers. The duration of word-internal vowels (i.e., reference) was compared with that of vowels at three different levels of prosodic boundaries: Word boundary, phonological phrase (PhP) boundary, and intonational phrase (IP) boundary. Note that the boundaries were coded a priori (i.e., based on the expectation that they should be located at specified positions in the reading texts). The analysis revealed that, while all learners showed evidence of some lengthening, their differentiation of boundary strength was poorer than that of native English speakers. For example, one of four beginning learners lengthened vowels at all three boundary conditions to an equal extent, while two other beginning learners showed similar lengthening effects at PhP- and IP-boundary conditions. The beginners showed a generally smaller magnitude of lengthening than native English speakers, which may result from the merging of higher-level prosodic constituents (e.g., IP) with lower-level ones (e.g., PhP and word). Advanced learners showed better differentiation of boundary strengths, but their magnitude of lengthening still fell slightly short of native English speech for all three conditions. This raises an additional possibility that the lengthening effect is overall smaller in Japanese than in English, which transferred to Japanese speakers' English and persistently remained even in the advanced learners. This relates to a recent finding that languages show different magnitudes of utterance-final lengthening, with English showing the strongest effect among nine other languages (Seifart et al., 2021). Taken together, the study by Ueyama (1996) suggests that Japanese learners of English show smaller degrees of phrase-final lengthening than native English speakers by the merging of higher- and lower-level prosodic constituents and/or the transfer of language-specific lengthening strength. It needs to be tested whether nativelike realization of L2 phrase-final lengthening is attainable.

1.4. Hypotheses

Based on the discussion above, several hypotheses can be formulated regarding how L1 Japanese speakers implement the spectral and temporal aspects of L2 AmE vowels. Note that the target learners of interest here are those who have learned the L2 for several years, similar

to those reported in Strange et al. (1998). The three L2 models' predictions diverge as to the learning outcome for such learners (i.e., which L2 categories would approach nativelike implementation and how likely, despite the expected influences from L1 categories), which the present study aims to test.

As for the spectral domain, the following hypotheses are proposed:

1. AmE tense-lax pairs /iː/-/ɪ/ and /uː/-/ʊ/ tend to be produced with overlapping qualities because they are categorized as Japanese /ii/-/i/ and /uu/-/u/, respectively. SLM(-r) predicts that 'dissimilar' /ɪ/ and /ʊ/ are more likely to form nativelike categories due to their lower categorical goodness than 'similar' /iː/ and /uː/, which should remain merged with the L1 categories and thus non-nativelike. PAM(-L2) predicts that perceptual learning is unlikely for these pairs, but it does not reject a possibility of category formation for phonetically 'deviant' /ɪ/ and /ʊ/. L2LP predicts that all the L2 categories can become equally nativelike in this SIMILAR scenario.
2. AmE /ɛ/ tends to be produced with a similar quality to Japanese /e/, which the L2 sound is categorized as. However, similar to the lax vowels above, all three models predict that nativelike quality can be achieved for this relatively 'dissimilar' or 'deviant' sound.
3. AmE /æː ʌ ɑː/ tend to be produced with overlapping qualities because they are categorized as Japanese /aa/ or /a/. Both SLM(-r) and PAM(-L2) predict that /æː/ is most likely to form a nativelike category due to its lowest category goodness, whereas the vowel is not special according to L2LP. Likewise, 'dissimilar' or 'deviant' /ʌ/ is more subject to category formation than 'similar' or 'better-fitting' /ɑː/ according to these two models, whereas such prediction is not made by L2LP. L2LP predicts that the copied L1 categories would need to be split to match the optimal L2 three-way contrast (NEW scenario), which is difficult yet achievable.

As for the temporal domain, the following hypotheses are proposed:

4. AmE phonetically long /iː æː ɑː uː/ tend to be produced with longer duration than AmE phonetically short /ɪ ɛ ʌ ʊ/ because they are categorized as Japanese phonologically long and short vowels, respectively. SLM(-r) and PAM(-L2) predict that the long-short distinction will diminish by new category formation for 'dissimilar' or 'deviant' short /ɪ ɛ ʌ ʊ/, whereby their very short duration owing to Japanese /i e a u/ becomes less short. L2LP also predicts that the L1-like stark durational distinction will diminish to match optimal L2 implementation, though to a lesser extent for /æː/-/ɛ/ and /ɑː/-/ʌ/ where duration serves as a secondary cue.
5. Stressed and unstressed vowels tend to be produced with an insufficient durational difference due to the transfer of Japanese mora-timed rhythm. Nativelike durational distinction can be achieved by shortening of unstressed vowels, lengthening of stressed ones, or both.

6. The effect of phrase-final lengthening tends to be insufficient because the levels of prosodic constituents are poorly differentiated and/or because the language-specific magnitude of lengthening is transferred. Nativelike lengthening may be achieved by establishing proper prosodic constituent levels and/or by overcoming the transfer.

These hypotheses are tested using the method explained in Section 2. The results are presented in Section 3 and then discussed in Section 4, followed by a brief conclusion in Section 5.

2. Method

The present study investigates speech data from the J-AESOP corpus (Kondo et al., 2015). The corpus has been developed as part of the Asian English Speech cOrpus Project (AESOP), a multi-national and multi-institutional project to construct L2 English speech corpora of Asian language speakers (Meng, Tseng, Kondo, Harrison, & Visceglia, 2009; Visceglia, Tseng, Kondo, & Sagisaka, 2009). The study utilizes a subset of the corpus, as explained below.

2.1. Speakers

J-AESOP features a total of 183 native Japanese speakers of English, all of whom were students at universities in the greater Tokyo area, Japan, where data collection took place. The present study restricts the sample to 102 speakers who had never resided outside of Japan (53 female, 49 male, mean age = 19.7) to control for their linguistic backgrounds. These speakers had received a formal English language education for six years in Japan (age 13–18), where reading and writing skills were of primary focus, but AmE was considered the norm in terms of pronunciation. They had also received some English instruction during college, the quality and quantity of which varied depending on the courses being offered. In addition to the Japanese speakers, the corpus also includes recordings of 13 native AmE speakers (nine female, four male, mean age = 21.5), which the present study also uses as a reference.

2.2. Materials

There are eight types of recording tasks in J-AESOP (see Kondo et al. (2015) for details). For the analysis of English vowels, the present study uses the data of Task 6_01, in which the participants read aloud "The North Wind and the Sun" (International Phonetic Association, 1999). The reading text contains all English vowel phonemes and therefore is suitable for the purpose of the present study. The full text is as follows (see Section 2.3 for the usage of the "¶" symbol):

> *The North Wind and the Sun were disputing which was the stronger, when a traveler came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveler take his cloak off should be considered stronger than the other. ¶ Then the North Wind blew as hard as he*

could, but the more he blew the more closely did the traveler fold his cloak around him; and at last the North Wind gave up the attempt. ¶ Then the Sun shone out warmly, and immediately the traveler took off his cloak. And so the North Wind was obliged to confess that the Sun was the stronger of the two.

The Japanese speakers also read aloud the Japanese version of "The North Wind and the Sun" (International Phonetic Association, 1999) in Task 6_02. The present study uses the data as well to compare the speakers' L1 and L2 productions. The romanized full text is as follows:

Aru toki, kitakaze to taiyou ga chikarakurabe o shimashita. Tabibito no gaitou o nugaseta hou ga kachi to iu koto ni kimete, mazu kitakaze kara hajimemashita. Kitakaze wa, 'Nani, hitomakuri ni shite miseyou' to, hageshiku fukitatemashita. Suruto tabibito wa, kitakaze ga fukeba fukuhodo gaitou o shikkari to karada ni kuttsukemashita. Kondo wa taiyou no ban ni narimashita. Taiyou wa kumo no aida kara, yasashii kao o dashite atatakana hikari o okurimashita. Tabibito wa, dandan yoi kokoromochi ni natte, shimai ni wa gaitou o nugimashita. Soko de, kitakaze no make ni narimashita.

2.3. Nativelikeness scores

As part of the development of the corpus, the Japanese speakers' spoken English levels were assessed by 16 phonetically trained judges with different L1s (four AmE, four Japanese, and eight other languages). Prior to the assessment, each speaker's recorded sample of Task 6_01 was divided into three sections (at the "¶" boundaries in Section 2.2). The judges listened to each section and evaluated the speech based on four criteria (segmental accuracy, prosody, fluency, and nativelikeness) on a 10-point scale each. The obtained scores were then averaged over the three sections, resulting in four kinds of scores for each speaker as assessed by a judge; the trisecting was to improve the accuracy of assessment by having the judges evaluate the same speaker three times. See Konishi (2022) for further details.

The present study uses the nativelikeness scores assessed by the four native AmE judges, where a value of one stands for "strongly foreign-accented" and 10 stands for "free of foreign accent." These scores are considered to reliably represent the speakers' nativelikeness (or lack of accentedness) in L2 English because native listeners, especially those with linguistic and pedagogical experience, are capable of consistently evaluating the phonological aspects of L2 speech, including accentedness (Saito, Trofimovich, & Isaacs, 2017). Inter-judge consistency was very high, with a Cronbach's alpha of 0.95 (cf. Saito et al., 2017). **Figure 2** shows the distribution of mean nativelikeness scores of the 102 Japanese speakers, averaged over the four judges. It can be seen that the speakers' nativelikeness scores vary substantially, ranging from 1.33 to 9.00 (mean = 4.74, median = 4.54, standard deviation = 1.51). The sample thus covers a large number of Japanese speakers with a wide range of perceived nativelikeness in L2 spoken English.

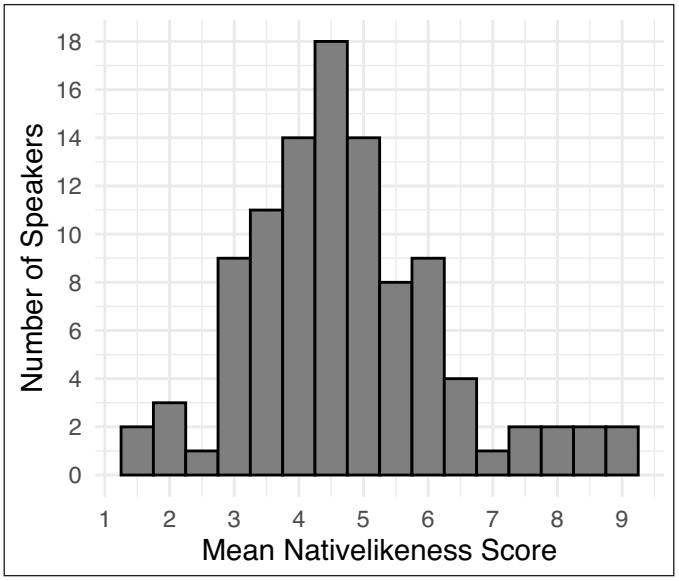

Figure 2: Distribution of mean nativelikeness scores.

2.4. Annotation

All recordings in J-AESOP have been annotated in Praat TextGrid format (Boersma & Weenink, 2023). The annotations of the English data are based on word- and segment-level forced alignment using the Hidden Markov Model Toolkit (Young et al., 2006) and the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Garofolo et al., 1993). As the transcription system of the TIMIT corpus assumes AmE pronunciations, the resulting segment labels are also based on AmE phonemes. As for Task 6_01, the output of automatic annotation has been further manually modified by trained phoneticians in the J-AESOP team to correct alignment errors. The annotators have also marked distinct speech events such as misreading (e.g., reading *cloak* as ***coat***), repetition (e.g., ***Then the Sun*** … *Then the Sun shone out warmly*), word insertion (e.g., *wrapped* ***around*** *in a warm cloak*), and disfluency (e.g., *should be* ***sh*** … *considered*) by assigning 'tags' to the relevant words (in bold above) so that these events can be distinguished from regular speech. Task 6_02 has also been annotated in the same procedure except that SPPAS (Bigi, 2015) was used for automatic alignment.

2.5. Data retrieval and additional coding

Based on the annotations, the Japanese speakers' production of 5985 English vowels (/iː/ = 983, /ɪ/ = 1766, /uː/ = 497, /ʊ/ = 304, /ɛ/ = 508, /æː/ = 706, /ʌ/ = 610, /ɑː/[5] = 611) in 37 kinds of words (**Table 2**) was retrieved from Task 6_01. The AmE speakers' production of 773 vowels was also retrieved (/iː/ = 129, /ɪ/ = 229, /uː/ = 65, /ʊ/ = 38, /ɛ/ = 65, /æː/ = 91, /ʌ/ = 78,

[5] TIMIT transcribes /ɑː/ and /ɔː/ somewhat interchangeably, which is reflected in the annotation of J-AESOP.

Vowel	Word
/iː/	*agr**ee**d, b**e**, close**ly**, h**e**, immedi**ate**l**y**, succ**ee**ded, warml**y***
/ɪ/	*cons**i**dered, d**i**d, d**i**sput**i**ng, h**i**m, h**i**s, **i**mmediately, **i**n, mak**i**ng, succeed**e**d, wh**i**ch, w**i**nd*
/uː/	*bl**ew**, disp**u**ting, tw**o**, wh**o***
/ʊ/	*c**ou**ld, sh**ou**ld, t**oo**k*
/ɛ/	*att**e**mpt, conf**e**ss, th**e**n, wh**e**n*
/æː/	***a**t, l**a**st, tr**a**veler, wr**a**pped*
/ʌ/	***o**ne, **o**ther, s**u**n, **u**p*
/ɑː/	*al**o**ng, **o**ff, str**o**nger*

Table 2: List of words containing target vowels (in bold).

/ɑː/ = 78). Vowels in tagged words were not included. Since the current version of J-AESOP does not contain prosodic annotations, the target vowels were further coded as either stressed (i.e., with primary or secondary stress) or unstressed (i.e., with no lexical stress) using the CMU Pronouncing Dictionary. Those in function words that are defined as stressed in the dictionary (e.g., /ɪ/ in "his") were coded as unstressed because they tend to become weak forms unless focused or pronounced in isolation (Selkirk, 2008). In addition, a vowel in the final syllable of a word that occurs at the end of a clause was coded as a clause-final vowel. These include: /iː/ in *agreed* and *warmly*; /ɪ/ in *disputing* and *him*; /ɛ/ in *attempt* and *confess*; /ɑː/ in *along* and *off* (in *take his cloak off*); /uː/ in *blew* (in *but the more he blew*) and *two*; and /ʊ/ in *could*. Relatedly, a vowel in the final syllable of a word that precedes a temporal pause was coded as an utterance-final vowel. Furthermore, postvocalic consonantal context was defined based on the segment after a target vowel to control for the effect of voicing on preceding vowel duration: Voiceless (/p t k f θ s ʃ t͡ʃ h/), voiced (/b d g v ð z ʒ d͡ʒ m n ŋ/), and others (e.g., approximants and vowels).[6]

The Japanese speakers' production of 18949 Japanese vowels (/i/ = 3724, /e/ = 2003, /a/ = 8757, /o/ = 3115, /u/ = 1350) was also retrieved from Task 6_02. Those in /ou/ were not included (cf. Note 2). Long vowels were not included either because they seldom appear in the text (e.g., /ii/ in *yasashii* 'gentle').

2.6. Acoustic measurement and normalization

For each vowel interval, the F1, F2, and duration were measured using Praat. The formants were the average values within the vowel interval, estimated by the built-in Burg algorithm with the formant ceiling setting of 5000 Hz for male speakers and 5500 Hz for female speakers.

[6] Vowel duration is shorter before voiceless than voiced consonants (often called voicing effect), which has been observed in both English (House, 1961; Mack, 1982) and Japanese (Yoneyama & Kitahara, 2014).

The obtained formant values were then Z-normalized per speaker, which is known as the Lobanov method (Lobanov 1971). Lobanov is a vowel-extrinsic (i.e., it uses information across multiple vowels) and formant-intrinsic (i.e., it operates on individual formants) method of formant normalization, which effectively eliminates spectral variation caused by physiological differences among speakers while preserving phonological and cross-linguistic differences (Adank, Smits, & van Hout, 2004). Vowel duration was also Z-normalized to control for speech rate while preserving the effects of linguistic factors such as intrinsic vowel duration, lexical stress, and phrase-final lengthening. Rate normalization is particularly important in L2 speech research because more proficient speakers tend to speak faster (Cucchiarini, Strik, & Boves, 2000; Yazawa & Kondo, 2022), making their vowel duration uniformly shorter and thus potentially obscuring the temporal characteristics of their speech. Note that the normalization was performed within Task 6_01.

2.7. Statistical analysis

In order to analyze the Japanese speakers' spectral and temporal implementation of AmE vowel categories in relation to their judged nativelikeness, a series of linear mixed effects (LME) models were fitted to the data of Task 6_01 using the lme4 (Bates, Mächler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) packages in R (R Core Team, 2021). A single model was fitted to each AmE vowel category (/iː ɪ uː ʊ ɛ æː ʌ ɑː/), yielding eight models in total. The function of each model was as follows:

$$lmer(score \sim F1.norm + F2.norm + dur.norm + (1|word) + (1|judge), data = target.category)$$

where *score* is the nativelikeness score, *F1.norm*, *F2.norm*, and *dur.norm* are the normalized F1, F2, and duration, *word* is the lexical item in which the vowel occurs, and *judge* is the evaluator of the nativelikeness score. The model thus evaluated how the normalized acoustic values of a particular vowel category would predict the speakers' overall perceived nativelikeness, taking inter-token and inter-judge variability into consideration.[7] *Speaker* was not included as a random effect because *score* (the response variable) is specific to each speaker; individual differences are not random variation to be controlled for but rather a factor to be investigated in the present study's design. The Z-normalization of the acoustic values can instead control for the individual variability that needs to be eliminated.

3. Results

Table 3 reports the results of the LME analyses, in which the estimated coefficients (β), standard errors, t and p values of the predictor variables are shown for each model (i.e., vowel category) per

[7] While a reversed model where nativelikeness predicts acoustic values is also possible, such modeling would result in a large number of models because there would be three response variables (F1, F2, and duration) for each vowel category and thus is less suitable.

		/iː/	/ɪ/	/uː/	/ʊ/	/ɛ/	/æː/	/ʌ/	/ɑː/
Intercept	*β*	3.991	5.039	4.422	4.865	4.663	4.528	5.940	3.573
	s.e.	0.563	0.554	0.572	0.559	0.563	0.642	0.588	0.578
	t	7.083	9.091	7.724	8.708	8.286	7.050	10.110	6.182
	p	0.004	0.003	0.003	0.003	0.003	<.001	<.001	0.005
F1	*β*	–0.463	0.078	–0.639	0.007	0.339	0.306	–0.303	0.641
	s.e.	0.069	0.036	0.097	0.125	0.085	0.076	0.070	0.066
	t	–6.675	2.156	–6.565	0.054	3.981	4.012	–4.325	9.765
	p	<.001	0.032	<.001	0.957	<.001	<.001	<.001	<.001
F2	*β*	0.457	–0.353	0.335	0.748	–0.864	0.280	0.632	–0.423
	s.e.	0.057	0.039	0.074	0.112	0.089	0.114	0.118	0.096
	t	8.047	–9.152	4.520	6.656	–9.689	2.463	5.374	–4.416
	p	<.001	<.001	<.001	<.001	<.001	0.014	<.001	<.001
Duration	*β*	–0.200	–0.101	0.135	0.132	–0.150	0.764	–0.274	0.181
	s.e.	0.028	0.030	0.037	0.054	0.047	0.053	0.046	0.045
	t	–7.252	–3.424	3.670	2.450	–3.178	14.359	–5.904	3.994
	p	<.001	<.001	<.001	0.038	0.002	<.001	<.001	<.001

Table 3: Results of LME models, where nativelikeness score is predicted by normalized acoustic values per vowel category.

column. For example, as for the production of /iː/, an increase of one standard deviation in F1 (since predictor variables were Z-normalized) predicts a decrease of 0.463 in the nativelikeness score from the estimated intercept 3.991; the effect of F1 is statistically significant ($t = –6.675, p < .001$) given the standard error of 0.069. Spectral and temporal results are reported separately below.

3.1. Spectral results

Figure 3 shows normalized F1 and F2 values of the Japanese speakers' AmE vowels in relation to their nativelikeness scores. Each circle shows mean formant values of a 0.50 score range (i.e., 1.25–1.75, 1.75–2.25 … and 8.75–9.25, as in the bins of **Figure 2**), where darker shades represent higher scores. The arrows point from lowest through intermediary to highest score ranges based on those mean values.[8] For comparison, mean formant values of the native AmE speakers' production in the corpus are also shown as diamonds.

[8] For /ɑː/, the lowest score range appears to be an outlier and therefore the second lowest range is used as the starting point.

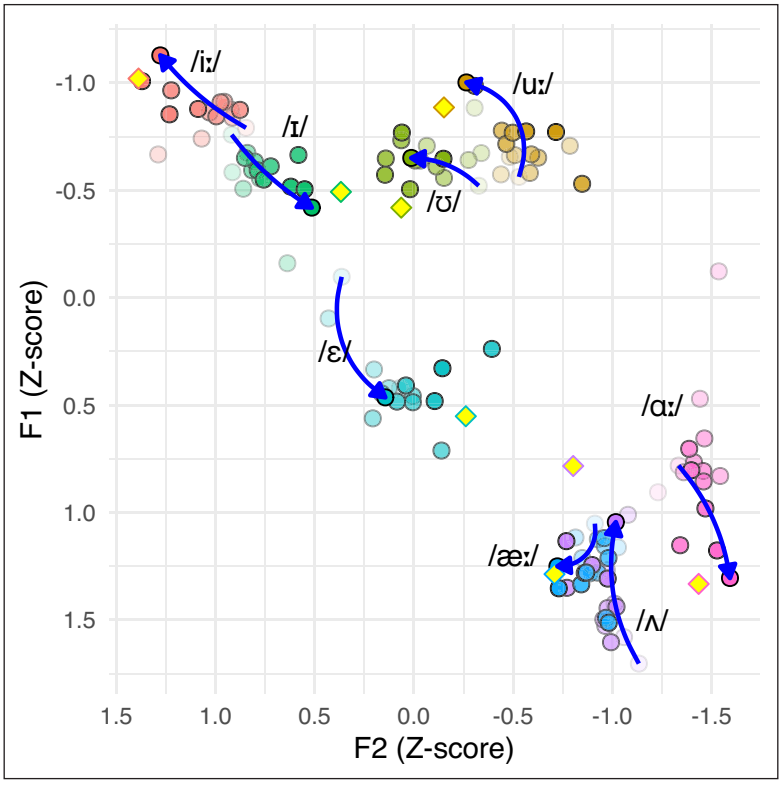

Figure 3: Normalized F1 and F2 of each vowel category according to nativelikeness score (diamonds = native AmE production).

Starting with /iː/ and /ɪ/, the figure shows nearly identical qualities of these vowels in the least nativelike speakers' production, whereas the qualities are farther apart from each other and closer to native AmE values in more nativelike speakers' production. The coefficients of the LME models suggest that a higher nativelikeness score is predicted by lower F1 and higher F2 for /iː/ and by higher F1 and lower F2 for /ɪ/. A similar tendency is found for /uː/ and /ʊ/, which showed overlapping qualities in less nativelike speakers' production but less overlapping and closer to native values in more nativelike speakers' production. The model coefficients suggest that those with higher nativelikeness scores produced /uː/ with lower F1 and higher F2 and /ʊ/ with higher F2. As for /ɛ/, both the figure and the model coefficients suggest that more nativelike speakers' production is characterized by higher F1 and lower F2 that are closer to native AmE values. Turning to /æː ʌ ɑː/, /ɑː/ is distinct from the other two vowels in showing much lower F2. More nativelike speakers produced the vowel with higher F1 and even lower F2 according to the coefficients. While /æː/ and /ʌ/ appear to be overlapping, it should be noted that the vowels show different qualities in the least nativelike speakers' production. More nativelike speakers produced /æː/ with higher F1 and higher F2 whereas /ʌ/ with lower F1 and higher F2 according to the coefficients, resulting in partial overlaps between the two vowels.

The qualities of /æː ʌ ɑː/, and /æː/ in particular, are close to native AmE values in the most nativelike speakers' production.[9]

To compare the Japanese speakers' AmE productions with their L1 categories, the F1 and F2 values of the AmE vowels were re-normalized with the Japanese data in the corpus. That is, the raw formant values of eight AmE monophthongs in Task 6_01 and five Japanese vowels in Task 6_02 were Z-normalized altogether per speaker so that the formant values are directly comparable across the two languages. **Figure 4** shows the outcome, in which the mean formant values of Japanese vowels are shown as black boxes. While the resultant Z-scores cannot be compared with those in **Figure 3**, the overall patterns of AmE spectral implementation remain largely the same. It can be seen in **Figure 4** that the least nativelike speakers' productions of AmE /iː ɪ/, /ɛ/, /æː ʌ/, /ɑː/, and /uː ʊ/ are generally proximal to Japanese /i/, /e/, /a/, /o/, /u/, respectively.

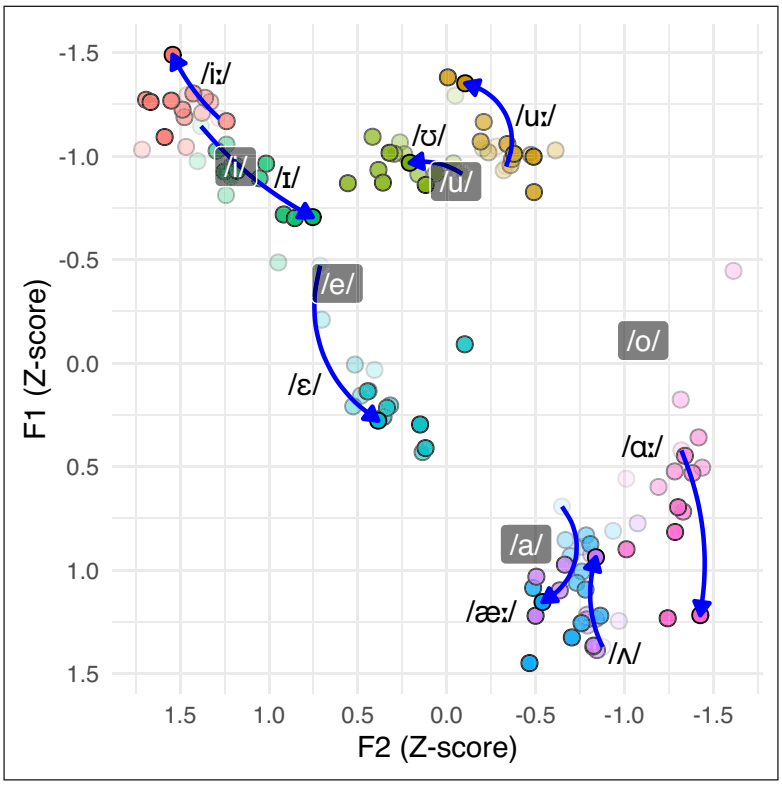

Figure 4: Re-normalized F1 and F2 of each vowel category according to nativelikeness score (black boxes = same speakers' L1 Japanese production).

[9] The F2 of /æː/ is likely undermeasured in the current data because the vowel is repeated four times in *traveler*, in which the preceding /tr/ causes tongue retraction (Deterding, 2006). Removing the word in fact yields much higher F2 values, especially in more nativelike speakers.

3.2. Temporal results

The LME models, which did not take prosodic factors into account, found that more nativelike speakers tended to produce phonetically long vowels /æː ɑː uː/ with longer duration than phonetically short vowels /ɪ ɛ ʌ/. Exceptions were /iː/ and /ʊ/, and word-by-word examinations of the data revealed that this was likely due to stress and phrase-final lengthening effects in the contexts in which the vowels occurred. Those with higher nativelikeness scores produced /iː/ in *be* and *he* with much shorter duration, which can be attributed to the proper shortening of unstressed vowels in function words. While the duration of /ʊ/ in *should* and *took* were similar across more and less nativelike speakers, more nativelike speakers produced /ʊ/ in *could* ("*Then the north wind blew as hard as he* ***could****, but …*") with substantially longer duration, presumably due to a stronger effect of phrase-final lengthening.

An additional LME model was thus fitted to further investigate the relationship between vowel duration and prosodic factors. The function of the model was as follows:

$$lmer(dur.norm \sim length + stress + clause + utterance + score : length + score : stress + score : clause + score : utterance + (1|judge) + (1|voicing), data = all.categories)$$

where *dur.norm* (normalized vowel duration) was set as the response variable this time. The predictor variable *length* is the phonetic length of AmE vowels (0 = short /ɪ ɛ ʌ ʊ/, 1 = long /iː æː ɑː uː/), which is hypothetically related to Japanese phonological long-short contrast (cf. **Table 1**). The other predictor variables were *stress* (0 = unstressed, 1 = stressed), *clause* (0 = clause-medial, 1 = clause-final), and *utterance* (0 = utterance-medial, 1 = utterance-final). The interactions of the four predictor variables with *score* were also included as predictors. The random effects were *judge* and *voicing* (postvocalic consonantal context),[10] while *speaker* was again not included. The result yielded significant main effects of *length* ($\beta = 0.271$, *s.e.* $= 0.030$, $t = 8.906$, $p < .001$) and *utterance* ($\beta = 0.706$, *s.e.* $= 0.040$, $t = 17.880$, $p < .001$). This indicates that phonetically long /iː æː ɑː uː/ were produced with longer duration than phonetically short /ɪ ɛ ʌ ʊ/ and vowel duration was longer utterance-finally than utterance-medially, in general. The main effects of *stress* ($\beta = 0.026$, *s.e.* $= 0.028$, $t = 0.934$, $p = .350$) and *clause* ($\beta = –0.015$, *s.e.* $= 0.045$, $t = –0.331$, $p = .741$) were not statistically significant. Moreover, the interaction of *score* was significant with *stress* ($\beta = 0.023$, *s.e.* $= 0.005$, $t = 4.221$, $p < .001$) and *clause* ($\beta = 0.064$, *s.e.* $= 0.009$, $t = 7.128$, $p < .001$). This suggests that more nativelike speakers exhibited more prominent effects of lexical stress and clausal boundary on vowel duration. The interaction of *score* was not significant with *length* ($\beta = 0.004$, *s.e.* $= 0.006$, $t = 0.643$, $p = .520$) and *utterance* ($\beta = –0.010$, *s.e.* $= 0.008$, $t = –1.301$, $p = .193$).

[10] The inclusion of *voicing* significantly improved the model fit according to a likelihood ratio test ($\chi^2(1) = 379$, $p < .001$).

Figure 5 shows the effects of *stress* and *clause* on normalized vowel duration in relation to the mean nativelikeness scores, with native AmE values shown as diamonds. The effect of *clause* is evident; the duration of clause-final vowels, both stressed and unstressed, is substantially longer in more nativelike speakers' production. As for clause-medial vowels, more nativelike speakers produced unstressed vowels with shorter duration and stressed ones with longer duration, though to a greater extent for the former. The most nativelike speakers' production is very close to native AmE vowels across all conditions.

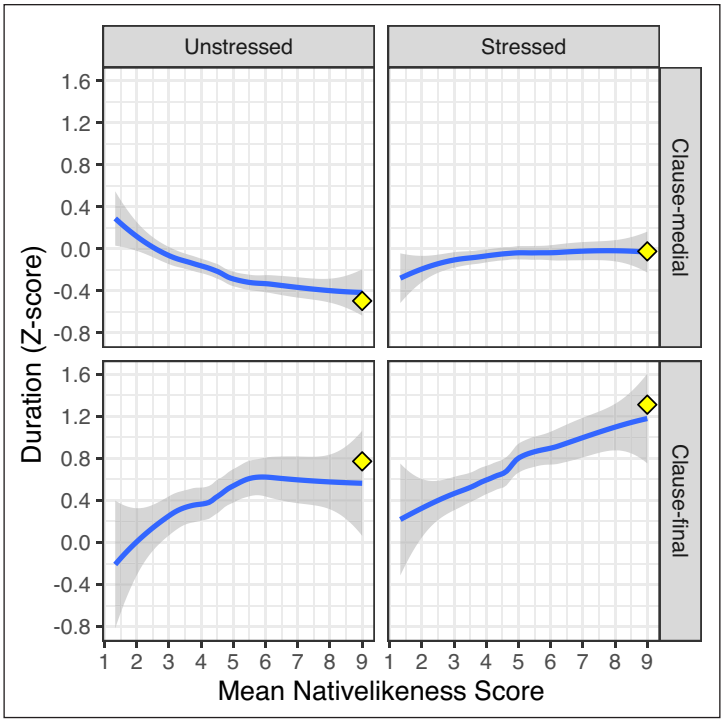

Figure 5: Mean normalized vowel duration (with 95% confidence intervals) in relation to nativelikeness score by stress and clausal position. Diamonds show mean native AmE productions.

4. Discussion

The analysis of the J-AESOP corpus revealed various patterns of spectral and temporal implementation of AmE vowels by Japanese speakers according to their judged L2 nativelikeness. The subsequent sections discuss theoretical interpretations of the results, followed by implications of the present study and directions for future research.

4.1. Spectral implementation

The spectral results were highly predictable from the acoustic and perceptual characteristics of L1 and L2 categories, in which less nativelike speakers' production was strongly influenced by L1

categories while more nativelike speakers' production was proximal to native AmE productions. However, there were some exceptions as discussed below.

The spectral qualities of the least nativelike speakers' vowels were generally predictable from the L1-L2 categorization patterns (cf. **Table 1**), largely supporting the first parts of the spectral hypotheses. As expected in Hypothesis 1, the qualities of AmE /iː/-/ɪ/ and /ʊ/-/ʊ/ were both overlapping and close to Japanese /i/ and /u/, respectively, suggesting that each L2 vowel pair was equated with, assimilated to, or copied from a single L1 vowel quality according to SLM(-r), PAM(-L2), and L2LP. Hypothesis 2 was also supported because AmE /ɛ/ was nearly identical to Japanese /e/ in quality, again conforming to the three models' predictions. Hypothesis 3, on the other hand, was only partially borne out. While the quality of AmE /æː/ was close to that of Japanese /a/, the quality of AmE /ʌ/ was somewhat distinct from either Japanese /a/ or native AmE /ʌ/. SLM(-r) is the only model that can explain this unexpected result. According to the model, it could be the case that a new category had been formed for the L2 sound due to its phonetic dissimilarity from the L1 sound, which was then 'deflected' away from the L1 category to maintain sufficient contrast in the common L1-L2 phonetic space. However, the obtained result is still at odds with the models' prediction that /æː/ should be more prone to category formation than /ʌ/.[11] It also remains unclear why category formation would not occur for the other lax vowels /ɪ ʊ ɛ/ that should be equally 'dissimilar' from L1 categories. As for AmE /ɑː/, the quality was closer to Japanese /o/ rather than the expected /a/. The obtained result is likely due to an orthographic influence because the AmE vowel is often spelled with "o" (e.g., *along*, *off*, and *stronger*), which represents /o/ in the Japanese romanization system. This account, while not explicitly stated in the hypotheses, is compatible with PAM(-L2)'s proposal that L1 and L2 categories with little phonetic similarity can still be assimilated at the phonological level (e.g., English /r/ ([ɹ]) and French /r/ ([ʁ])).

More nativelike speakers' production was characterized by closer proximity to native AmE targets, with the most nativelike speakers' production being very close to native productions, across all eight AmE vowel categories (similar to Oh et al. (2011)'s finding for children). The result is consistent with Hypothesis 2, which predicted that Japanese speakers can achieve nativelike quality of AmE /ɛ/ by creating a new category for the 'dissimilar' or 'deviant' L2 sound according to SLM(-r) and PAM(-L2), or by modifying the copy of Japanese /e/ according to L2LP. Hypothesis 3 was also consistent with the result. SLM(-r) and PAM(-L2) predicted that new category formation is very likely for AmE /æː/, which aligns with the observed very nativelike quality of the L2 vowel. The quality of AmE /ʌ/ may be related to the above category formation,

[11] An alternative possibility is that the F1 of Japanese /a/ is undermeasured in the current data, since Nishi et al.'s (2008) study with a stricter control of phonetic context found much higher F1 values for the vowel (cf. Figure 1). If so, then it is /æː/ that has undergone category formation instead of /ʌ/, which would align well with the predictions of SLM(-r) and PAM(-L2).

as the vowel seems to maintain phonetic contrast from AmE /æː/ as proposed in SLM (-r). Category formation can also explain the nativelike quality of AmE /ɑː/, as the vowel should be phonetically dissimilable from Japanese /o/ as proposed in PAM(-L2). L2LP's prediction that AmE /æː ʌ ɑː/ would all approach and reach optimal L2 implementation was also supported. However, regarding Hypothesis 1, the prediction of SLM(-r) that 'dissimilar' /ɪ ʊ/ are more likely to show nativelike qualities than 'similar' /iː uː/ does not seem to be supported. More nativelike speakers produced /iː/ with more peripheral qualities simultaneously as they produced /ɪ/ with more centralized qualities, which does not accord with the prediction that 'similar' /iː/ will remain linked to the Japanese /i/ quality. Also, more nativelike speakers seem to have raised /uː/ to establish a qualitative contrast from /ʊ/, which contradicts the prediction that 'similar' /uː/ will remain merged with the Japanese /u/ quality. 'Deflection' from L1 categories does not explain either result because these 'similar' L2 categories (/iː uː/) are not subject to category formation in the first place. Also unsupported was the prediction of PAM(-L2) that learning is unlikely to occur for these tense-lax pairs and, even if it occurs, only 'deviant' /ɪ ʊ/ would undergo category formation. The obtained results are more compatible with L2LP's explanation that copied L1 categories gradually approach optimal L2 implementation regardless of L1-L2 similarities (cf. Yazawa et al., 2020).

Taken together, the spectral results provide mixed support for the predictions of the three L2 models. SLM(-r) is the only model that can explain the distinct quality of AmE /ʌ/ (i.e., 'deflection' from L1 category), but the model's claim that 'similar' L2 sounds are likely to remain non-nativelike is questionable based on the current results. Likewise, PAM(-L2) has strength in explaining the Japanese /o/-like quality of AmE /ɑː/ (phonological assimilation despite phonetic dissimilarity), but its prediction that good discrimination hinders successful categorical learning was unsupported. The discrepancies may derive from the sheer difficulty of defining and measuring cross-linguistic similarities. Flege (1995, p. 264) noted that an objective means for gauging the degree of perceived cross-linguistic phonetic distance is yet to be defined, which Flege and Bohn (2021, p. 33) re-acknowledged recently. Best and Tyler (2007, p. 26) also noted that how listeners identify nonnative phones as equivalent to L1 phones has not received adequate treatment in any L2 model. Relatedly, it could be the case that acoustic cues that were not examined in the study, such as vowel inherent spectral change (VISC), voice quality, and pitch are necessary to better characterize L1-L2 similarities. VISC seems a particularly relevant aspect for L2 English vowel production (Schwartz & Kaźmierski, 2020), though the present study did not probe into it due to its complexity and in the interests of space (but see Yazawa (2022)). Turning to L2LP, while the model's prediction that all L2 categories can become optimal was upheld, it must be noted that a sizable number of speakers did not arrive at optimal L2 implementation. The model thus may be overpredicting success, though it remains to be seen whether the learners will eventually acquire

nativelike productions. It is also unclear why non-optimal AmE /ʌ/ would emerge in the 'copy' of Japanese vowel categories that lacks the vowel in the least nativelike speakers' production. The model proposes that L2 learners can split existing categories to form new ones in the L2 grammar (FAMILIAR NEW scenario), but the exact mechanism is somewhat underspecified and needs more elaboration, as noted by Escudero (2005, p. 317). Perhaps feature-based modeling can be useful for further investigating the mechanism, as discussed by Escudero and Boersma (2004), and Yazawa (2020).

4.2. Temporal implementation

The temporal implementation also showed evidence of L1-L2 categorical relationship. In accordance with the first part of Hypothesis 4, the Japanese speakers produced AmE long /iː æː ɑː uː/ with significantly longer duration than AmE short /ɪ ɛ ʌ ʊ/, presumably because these vowels are assimilated to Japanese phonologically long and short vowels, respectively. However, the three models' prediction that the length distinction would be less prominent in more nativelike speakers was not supported, as no significant interaction was found between phonetic length and nativelikeness score. Moreover, the coefficients of the LME models in **Table 3** suggest that those with higher nativelikeness scores produced AmE long /æː/ and /ɑː/ with even longer duration and AmE short /ɛ/ and /ʌ/ with even shorter duration, thus enhancing the durational difference. It is possible that more nativelike speakers have done so to make these vowels maximally distinct because the vowels show a great degree of spectral overlap and thus are difficult to differentiate solely by vowel spectra (Hillenbrand et al., 2000). This explanation is in partial agreement with L2LP, which predicted that the length distinction should remain less changed for /æː/-/ɛ/ and /ɑː/-/ʌ/ for which duration serves an informative secondary cue. However, the above interpretation is speculative because the coefficients in **Table 3** are most likely mediated by prosodic factors such as lexical stress and phrase-final lengthening, as discussed below.

The statistical analysis found a significant interaction between stress and nativelikeness score. In support of Hypothesis 5, those with higher nativelikeness scores showed greater durational differences between stressed and unstressed vowels, suggesting their proper manifestation of English stress rhythm. Judging from **Figure 5**, nativelikeness seems to be predicted by both shortening of unstressed vowels and lengthening of stressed ones, though the pattern is not symmetric. Whereas the least nativelike speakers underutilized duration for both stressed and unstressed vowels, average Japanese learners of English (e.g., mean nativelikeness score = 4.5) showed nativelike duration for stressed vowels but not for unstressed ones. This suggests that the shortening of unstressed vowels is more difficult to acquire than the lengthening of stressed ones, adding to the previous findings (Kondo, 2009; Lee et al., 2006). The most nativelike speakers

showed nativelike implementation of unstressed vowels as well, which agrees with Lee et al. (2006)'s proposal that Japanese speakers can acquire nativelike shortening of English unstressed vowels because of their increased sensitivity to L1 phonemic vowel length contrasts. If this is the case, then a phonological feature shared by multiple L1 categories such as length transfers to L2 prosodic acquisition, a possibility that current L2 models may need to additionally consider.

Turning to the effect of phrase-final lengthening, while clause-final vowels were not significantly longer than clause-medial ones per se, the interaction between clause boundary and nativelikeness score was significant. The result, together with **Figure 5**, aligns with the prediction in Hypothesis 6 that the effect of phrase-final lengthening is generally insufficient in the learners' speech but can approach nativelike implementation, though it remains unclear why. It is possible that a clause-final position, where an IP boundary is expected to fall (Selkirk, 2009), received a boundary of another, interlanguage prosodic constituent (e.g., IP-PhP compound) in less nativelike speech, as predicted in the hypothesis. The transfer of language-specific magnitude of lengthening is less feasible, as the speakers showed comparable lengthening in utterance-final positions (i.e., for the same prosodic constituent level) regardless of their nativelikeness.[12] However, it must be kept in mind that the above analysis is preliminary at best due to methodological limitations. For one thing, the assumption that a syntactic clause boundary should correspond to an IP boundary may not necessarily hold in learner speech, which tends to show different prosodic phrasal parsing from native speech (Jun & Oh, 2000). For another, a temporal pause used to identify an utterance may not reliably indicate the end of an actual utterance. For a more fully-fledged analysis of phrase-final lengthening, a more proper prosodic annotation of the data is certainly needed. Perhaps the framework of Tones and Break Indices (ToBI; Jun, 2005) can be useful. However, since each ToBI system is specific to each language (e.g., MAE-ToBI for Mainstream American English (Beckman, Hirschberg, & Shattuck-Hufnagel, 2005) and J_ToBI for Tokyo Japanese (Venditti, 2005)), it needs to be tested whether and how these systems can be applied to L1-L2 interlanguage prosody. For example, a 'pitch reset' is an obligatory indicator of an IP boundary in J_ToBI but not in MAE-ToBI (Igarashi, Nishikawa, Tanaka, & Mazuka, 2013). If what is seemingly a pitch reset is observed in Japanese speakers' English, should it indicate an IP boundary or not? More fundamentally, can the IP in J_ToBI and that in MAE-ToBI be equated in the first place? Many similar questions arise, which must be carefully considered before the ToBI system can be reliably applied to L2 speech data.

In summary, while the relationship between L1 phonologically long and short categories and L2 phonetically long and short categories did shape the Japanese speakers' temporal

[12] When vowel duration was re-normalized across Tasks 6_01 and 6_02, utterance-final vowel duration was substantially shorter in Japanese than in English. It thus remains inconclusive whether the transfer account should be abandoned.

implementation of AmE vowels, prosodic factors such as stress and phrase-final lengthening turned out to be more important predictors of the speakers' nativelikeness insofar as vowel duration is concerned. Although current L2 models primarily focus on segmental categories, prosodic factors are also necessary to fully account for L2 speech acquisition. There have been a few attempts to expand the models' predictions to prosodic factors such as lexical tone (Chen, Best, & Antoniou, 2020; Escudero, Smit, & Mulak, 2022; Hao, 2014), which should be further pursued. The application of the ToBI framework to L2 speech, though having a long way to go, can also open up new avenues for future research.

4.3. Implications and future directions

The present study has demonstrated that a corpus-based approach can be used not only to replicate the results of previous studies but also to complement those results with new insights. Such corpus-based studies of L2 speech, though currently few, have a large potential in the era of online collaborative research and big data. An important avenue for extending the present study is to further examine the Japanese speakers' production of L1 vowel categories in relation to their L2 nativelikeness using the same corpus, which is expected to answer a theoretically relevant question: Do acquired L2 categories in turn affect L1 categories? Unlike L2LP which assumes separate L1 and L2 grammars and no direct influence of L1 categories on L2 ones, SLM(-r) and PAM(-L2) both propose that L1 and L2 categories exist in a common space and can affect each other in a bidirectional manner (i.e., L1 affecting L2 (forward transfer) and L2 affecting L1 (backward transfer)). A growing body of research suggests that late L2 learners, highly proficient ones in particular, can show lasting changes in L1 pronunciation presumably due to assimilation of L1 categories to L2 ones (Kartushina, Frauenfelder, & Golestani, 2016). However, the relationship between the magnitude of backward transfer and L2 proficiency is not entirely clear. In particular, it needs to be addressed whether late learners who are highly proficient but not dominant in the L2 would show evidence of backward transfer, a phenomenon often associated with L2 dominance and L1 attrition. The speakers investigated in the present study are a good representative of such a population. While it has been assumed that their L1 categories are invariant, slight differences may exist between more and less nativelike speakers' implementation (Yazawa, 2021). Further investigation of J-AESOP can help shed more light on the dynamic relationship between L1 and L2 categories.

Finally, while the present study has sought theoretical explanations of L1 Japanese speakers' L2 AmE vowel production, the results have important pedagogical implications as well. Despite their relatively similar linguistic background, the Japanese speakers exhibited diverse levels of perceived nativelikeness in L2 AmE, ranging from heavily foreign-accented to practically nativelike. This diversity can be attributed to various inter- and intra-learner factors, including the quality and quantity of input, teaching and learning strategies and styles, individuals' auditory acuity,

working memory, and motivation, among others. Although it would be impossible to investigate all of these factors, the fact that late learners of English in classroom settings without any overseas experience could achieve near-nativelike production is simply encouraging. Teachers of English may also find the result of the present study relevant and applicable. For example, **Figure 4** shows the correspondence between vowel quality and perceived nativelikeness in Japanese speakers' English in relation to their L1 categories, which could be useful for making their pronunciations less accented and more intelligible. In addition, **Figure 5** suggests that the reduction of unstressed vowels should be emphasized in teaching average Japanese learners of English and that phrasal prosody deserves more attention as it is currently seldom taught. Thus, the development and research of L2 speech corpora are beneficial for pedagogical purposes as well.[13]

5. Conclusion

The present study has provided a comprehensive picture of how adult Japanese speakers implement L2 AmE vowel categories in relation to their L1 categories, adding to the previous literature by adopting a corpus-based approach. The spectral implementation was highly predictable from the characteristics of L1 and L2 categories; less nativelike speakers' production was strongly influenced by L1 categories, except AmE /ʌ/ that could be deflected away from Japanese /a/ according to SLM(-r) and AmE /ɑː/ that seemed orthographically assimilated to Japanese /o/ according to PAM(-L2), while more nativelike speakers' production was proximal to target L2 categories for all vowels, in accordance with L2LP. In contrast, the temporal implementation was characterized better by how well the prosodic factors such as stress and phrase-final lengthening were realized. This indicates that the L2 models, which are currently centered around segmental categories, should be extended to incorporate the prosodic level as well. These results demonstrate the usefulness of bilingual speech corpora such as J-AESOP for investigating the interactions between L1 and L2 categories, which is expected to provide useful new insights into theories of L2 speech acquisition and their application.

[13] Theoretical models of L2 speech acquisition can also have their own implications for language learning and teaching. See Elvin and Escudero (2019) for the case of L2LP.

Acknowledgements

The authors would like to thank Guest Editor Cinzia Avesani and two anonymous reviewers for their very helpful and constructive feedback, which greatly improved the quality and readability of the manuscript.

Competing interests

The authors have no competing interests to declare.

Author contributions

Kakeru Yazawa devised the conception and design of the study and led the drafting and revision of the manuscript. Takayuki Konishi collected the nativelikeness scores and drafted parts of the manuscript. James Whang carried out the statistical analyses and drafted parts of the manuscript. Paola Escudero contributed to the theoretical predictions and interpretations of the results and commented on different versions of the manuscript. Mariko Kondo supervised the work as the chief administrator of the J-AESOP corpus.

References

Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, *116*(5), 3099–3107. DOI: https://doi.org/10.1121/1.1795335

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42*(4), 529–555. DOI: https://doi.org/10.1111/j.1467-1770.1992.tb01043.x

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixedeffects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 9–54). Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199249633.003.0002

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonation structure in Japanese and English. *Phonology Yearbook*, *3*(1), 255–309. DOI: https://doi.org/10.1093/acprof:oso/9780199249633.003.0002

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.

Best, C. T., & Tyler, M. (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* (pp. 13–34). Amsterdam/Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/lllt.17.07bes

Bigi, B. (2015). SPPAS – Multi-lingual approaches to the automatic annotation of speech. *The Phonetician, 111–112*, 54–69.

Boersma, P., & Weenink, D. (2023). *Praat: doing phonetics by computer (Version 6.3.03)*. Retrieved from http://www.praat.org/

Chen, J., Best, C. T., & Antoniou, M. (2020). Native phonological and phonetic influences in perceptual assimilation of monosyllabic Thai lexical tones by Mandarin and Vietnamese listeners. *Journal of Phonetics, 83*, 101013. DOI: https://doi.org/10.1016/j.wocn.2020.101013

Chen, J., & Chang, H. (2022). Sketching the landscape of speech perception research (2000–2020): A bibliometric study. *Frontiers in Psychology*, *13*, 822241. DOI: https://doi.org/10.3389/fpsyg.2022.822241

Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *The Journal of the Acoustical Society of America, 107*(2), 989–999. DOI: https://doi.org/10.1121/1.428279

Deterding, D. (2006). The North Wind versus a Wolf: Short texts for the description and measurement of English pronunciation. *Journal of the International Phonetic Association, 36*(2), 187–196. DOI: https://doi.org/10.1017/S0025100306002544

Elvin, J., & Escudero, P. (2019). Cross-linguistic influence in second language speech: Implications for learning and teaching. In M. J. Gutierrez-Mangado, M. Martínez-Adrián & F. Gallardo-del-Puerto (Eds.), *Cross-linguistic influence: From empirical evidence to classroom practice* (pp. 1–20). Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-22066-2_1

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization* (Doctoral dissertation). Utrecht University.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition, 26*(4), 551–585. DOI: https://doi.org/10.1017/S0272263104040021

Escudero, P., Smit, E. A., & Mulak, K. E. (2022). Explaining L2 lexical learning in multiple scenarios: Cross-situational word learning in L1 Mandarin L2 English speakers. *Brain Sciences*, *12*(12), 1618. DOI: https://doi.org/10.3390/brainsci12121618

Escudero, P., & Yazawa, K. (in press). The Second Language Linguistic Perception Model (L2LP). In M. Amengual (Ed.), *The Cambridge handbook of bilingual phonetics and phonology*. Cambridge University Press.

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *The Journal of the Acoustical Society of America, 97*(3), 1893–1904. DOI: https://doi.org/10.1121/1.412063

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

Flege, J. E., & Bohn, O.-S. (2021). The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/9781108886901.002

Fox, R. A., & McGory, J. T. (2007). Second language acquisition of a regional dialect of American English by native Japanese speakers. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 117–134). Amsterdam/ Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/lllt.17.13fox

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, *27*(4), 765–768. DOI: https://doi.org/10.1121/1.1908022

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., Dahlgren, N. L., & Zue, V. (1993). *TIMIT Acoustic Phonetic Continuous Speech Corpus*. Philadelphia: Linguistic Data Consortium. Retrieved from https://catalog.ldc.upenn.edu/ldc93s1

Gay, T. (1978). Physiological and acoustic correlates of perceived stress. *Language and Speech*, *21*(4), 347–353. DOI: https://doi.org/10.1177/002383097802100409

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 515–546). Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110197105.2.515

Hao, Y.-C. (2014). The application of the Speech Learning Model to the L2 acquisition of Mandarin tones. In *Proceedings of 4th International Symposium on Tonal Aspects of Languages* (pp. 67–70). Nijmegen: International Speech Communication Association.

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, *108*(6), 3013–3022. DOI: https://doi.org/10.1121/1.1323463

Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111. DOI: https://doi.org/10.1121/1.411872

Hirata, Y. (2004). Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics*, *32*(4), 565–589. DOI: https://doi.org/10.1016/j.wocn.2004.02.004

House, A. S. (1961). On vowel duration in English. *The Journal of the Acoustical Society of America*, *33*(9), 1174–1178. DOI: https://doi.org/10.1121/1.1908941

Igarashi, Y., Nishikawa, K., Tanaka, K., & Mazuka, R. (2013). Phonological theory informs the analysis of intonational exaggeration in Japanese infant-directed speech. *The Journal of the Acoustical Society of America*, *134*(2), 1283–1294. DOI: https://doi.org/10.1121/1.4812755

International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to use the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Jun, S.-A. (2005). *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199249633.001.0001

Jun, S.-A., & Kubozono, H. (2020). Asian Pacific rim. In *The Oxford handbook of language prosody* (pp. 355–369). Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780198832232.013.23

Jun, S.-A., & Oh, M. (2000). Acquisition of second language intonation. In *Proceedings of the 6th international conference on spoken language processing*, *4*, 76–79. DOI: https://doi.org/10.21437/ICSLP.2000-754

Kartushina, N., Frauenfelder, U. H., & Golestani, N. (2016). How and when does the second language influence the production of native speech sounds: A literature review. *Language Learning*, *66*(S2), 155–186. DOI: https://doi.org/10.1111/lang.12187

Keating, P. A., & Huffman, M. K. (1984). Vowel variation in Japanese. *Phonetica*, *41*(4), 191–207. DOI: https://doi.org/10.1159/000261726

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: fundamental frequency lends little. *The Journal of the Acoustical Society of America*, *118*(2), 1038–1054. DOI: https://doi.org/10.1121/1.1923349

Kondo, M. (2009). Is acquisition of L2 phonemes difficult? Production of English stress by Japanese speakers. In *Proceedings of the 10th generative approaches to second language acquisition conference* (pp. 105–112). Somerville, MA: Cascadilla Proceedings Project.

Kondo, M., Tsubaki, H., & Sagisaka, Y. (2015). Segmental variation of Japanese speakers' English: Analysis of "the North Wind and the Sun" in AESOP corpus. *Journal of the Phonetic Society of Japan*, *19*(1), 3–17. DOI: https://doi.org/10.24467/onseikenkyu.19.1_3

Konishi, T. (2022). *A corpus-based study on Japanese English rhythm* (PhD thesis, Waseda University). Retrieved from http://hdl.handle.net/2065/00088943

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. DOI: https://doi.org/10.18637/jss.v082.i13

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change*. Walter De Gruyter. DOI: https://doi.org/10.1515/9783110167467

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*, 227–249. DOI: https://doi.org/10.1017/S0142716405050150

Lee, B., Guion, S. G., & Harada, T. (2006). Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition*, *28*(3), 487–513. DOI: https://doi.org/10.1017/S0272263106060207

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, *49*(2B), 606–608. DOI: https://doi.org/10.1121/1.1912396

Mack, M. (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *The Journal of the Acoustical Society of America*, *71*(1), 173–178. DOI: https://doi.org/10.1121/1.387344

Meng, H., Tseng, C.-y., Kondo, M., Harrison, A., & Visceglia, T. (2009). Studying L2 suprasegmental features in Asian Englishes: A position paper. In *Proceedings of the 10th annual conference of the international speech communication association* (pp. 1715–1718). Brighton: International Speech Communication Association. DOI: https://doi.org/10.21437/Interspeech.2009-517

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73–97. DOI: https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R., & Trent-Brown, S. A. (2008). Acoustic and perceptual similarity of Japanese and American English vowels. *The Journal of the Acoustical Society of America*, *124*(1), 576–588. DOI: https://doi.org/10.1121/1.2931949

Nogita, A., Yamane, N., & Bird, S. (2013). The Japanese unrounded back vowel /ɯ/ is in fact unrounded central/front [ʉ] – [ʏ]. In *Ultrafest VI program and abstract booklet* (pp. 39–42).

Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (2011). A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. *Journal of Phonetics*, *39*(2), 156–167. DOI: https://doi.org/10.1016/j.wocn.2011.01.002

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America*, *32*(6), 693–703. DOI: https://doi.org/10.1121/1.1908183

Port, R. F., Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in Japanese. *The Journal of the Acoustical Society of America*, *81*(5), 1574–1585. DOI: https://doi.org/10.1121/1.394510

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria. Retrieved from https://www.r-project.org/

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292. DOI: https://doi.org/10.1016/S0010-0277(00)00101-3

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*(4), 439–462. DOI: https://doi.org/10.1093/applin/amv047

Schwartz, G., & Kaźmierski, K. (2020). Vowel dynamics in the acquisition of L2 English – an acoustic study of L1 Polish learners. *Language Acquisition*, *27*(3), 227–254. DOI: https://doi.org/10.1080/10489223.2019.1707204

Seifart, F., Strunk, J., Danielsen, S., Hartmann, I., Pakendorf, B., Wichmann, S., … Bickel, B. (2021). The extent and degree of utterance-final word lengthening in spontaneous speech from 10 languages. *Linguistics Vanguard*, *7*(1). DOI: https://doi.org/10.1515/lingvan-2019-0063

Selkirk, E. (2008). The prosodic structure of function words. In J. J. McCarthy (Ed.), *Optimality theory in phonology: A reader* (pp. 464–482). Oxford: Blackwell. DOI: https://doi.org/10.1002/9780470756171.ch25

Selkirk, E. (2009). On clause and intonational phrase in Japanese: The syntactic grounding of prosodic constituent structure. *GENGO KENKYU*, *136*, 35–73. DOI: https://doi.org/10.11435/gengo.136.0_35

Strange, W., Akahane-Yamada, R., Kubo, R., Trent-Brown, S. A., Nishi, K., & Jenkins, J. J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, *26*(4), 311–344. DOI: https://doi.org/10.1006/jpho.1998.0078

Sugimoto, J., & Uchida, Y. (2020). English phonetics and teacher training: Designing a phonetics course for japanese preservice teachers. *Journal of the Phonetic Society of Japan*, *24*, 22–35. DOI: https://doi.org/10.24467/onseikenkyu.24.0_22

Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, *35*(4), 445–472. DOI: https://doi.org/10.1016/j.wocn.2006.12.001

Ueyama, M. (1996). Phrase-final lengthening and stress-timed shortening in the speech of native speakers and Japanese learners of English. In H. T. Bunnell & W. Idsardi (Eds.), *Proceeding of the 4th international conference on spoken language processing*, *2*, 610–613. Philadelphia: Institute of Electrical and Electronics Engineers. DOI: https://doi.org/10.1109/ICSLP.1996.607435

Umeda, N. (1975). Vowel duration in American English. *The Journal of the Acoustical Society of America*, *58*(2), 434–445. DOI: https://doi.org/10.1121/1.380688

van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, *6*, 1000. DOI: https://doi.org/10.3389/fpsyg.2015.01000

Venditti, J. J. (2005). The J_ToBI model of Japanese intonation. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 172–200). Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199249633.003.0007

Visceglia, T., Tseng, C.-y., Kondo, M., & Sagisaka, Y. (2009). Phonetic aspects of content design in AESOP (Asian English Speech cOrpus Project). In *Proceedigns of 2009 Oriental COCOSDA international conference on speech database and assessments* (pp. 52–57). Beijing: Institute of Electrical and Electronics Engineers. DOI: https://doi.org/10.1109/ICSDA.2009.5278376

Warner, N., & Arai, T. (2001). Japanese mora-timing: A review. *Phonetica*, *58*(1–2), 1–25. DOI: https://doi.org/10.1159/000028486

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, *91*(3), 1707–1717. DOI: https://doi.org/10.1121/1.402450

Yazawa, K. (2020). *Testing Second Language Linguistic Perception : A case study of Japanese, American English, and Australian English vowels* (Doctoral dissertation, Waseda University, Tokyo). Retrieved from http://hdl.handle.net/2065/00073679

Yazawa, K. (2021). L1 phonetic drift in vowel production by Japanese learners of English. In *Proceedings of the 35th General Meeting of the Phonetic Society of Japan* (pp. 113–118).

Yazawa, K. (2022). Transfer of incomplete neutralization: A case of /ei/ and /ou/ in Japanese. *Second Language*, *20*, 47–59. DOI: https://doi.org/10.11431/secondlanguage.20.0_47

Yazawa, K., & Kondo, M. (2019). Acoustic characteristics of Japanese short and long vowels: Formant displacement effect revisited. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (Eds.), *Proceedings of the 19th international congress of phonetic sciences* (p. 100). Canberra: Australasian Speech Science and Technology Association Inc.

Yazawa, K., & Kondo, M. (2022). A comparison of rhythm metrics for L2 speech. In *Proceedings of the 11th International Conference on Speech Prosody* (pp. 332–336). International Speech Communication Association. DOI: https://doi.org/10.21437/SpeechProsody.2022-68

Yazawa, K., Ozaki, Y., Short, G., Kondo, M., & Sagisaka, Y. (2015). A study of the production of unstressed vowels by Japanese speakers of English using the J-AESOP corpus. In *Proceedings of the 18th Oriental COCOSDA/CASLRE* (pp. 96–100). Shanghai: Institute of Electrical and Electronics Engineers. DOI: https://doi.org/10.1109/ICSDA.2015.7357872

Yazawa, K., Whang, J., Kondo, M., & Escudero, P. (2020). Language-dependent cue weighting: An investigation of perception modes in L2 learning. *Second Language Research, 36*(4), 557–581. DOI: https://doi.org/10.1177/0267658319832645

Yoneyama, K., & Kitahara, M. (2014). Voicing effect on vowel duration: Corpus analyses of Japanese infants and adults, and production data of English learners. *Journal of the Phonetic Society of Japan, 18*(1), 30–39. DOI: https://doi.org/10.24467/onseikenkyu.18.1_30

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., … Woodland, P. (2006). *The HTK Book (for HTK Version 3.4)*. Cambridge: Cambridge University Engineering Department. Retrieved from http://htk.eng.cam.ac.uk