

# The Relative Contributions of Duration and Amplitude to the Perception of Japanese-accented English as a Function of L2 Experience

Saya Kawase, Jeesun Kim, Chris Davis

The MARCS Institute, Western Sydney University, Sydney, NSW, Australia

s.kawase@westernsydney.edu.au

## Abstract

This study investigated the effect of duration and amplitude of speech produced by non-native (L2) talkers on foreign accent and comprehensibility ratings. We selected two Japanese L2 talkers with different experience in producing L2 English and then transplanted their speech duration and amplitude values onto English native speech. Native English listeners rated the transplanted sentences as well as untransplanted ones (control). Transplanted sentences were rated as more accented and more difficult to understand than control sentences. The degree of L2 experience only made a difference for the perceived degree of foreign accent ratings when duration was transplanted.

**Index Terms:** foreign-accented speech perception, speech rhythm, second language speech

## 1. Introduction

Speech produced by second language (L2) talkers is often perceived as having a ‘foreign accent’. A foreign accent typically occurs due to the influence of a talker’s first language (L1), namely differences between L1 and L2 at segmental and/or supra-segmental levels. For example, in terms of segments, an accented L2 talker may miss, insert or substitute phones. At the supra-segmental level, an L2 talker’s speech rhythm, intonation or prosody may be different.

Foreign-accented speech is typically less intelligible and/or comprehensible than native speech and requires more listening effort [1]. Interestingly, the degree of perceived accent does not necessarily correspond to the degree of recognition difficulties. For example, listeners may be able to correctly identify words spoken by L2 talkers and yet perceive these as ‘heavily accented’ [2]. Likewise, the perceived degree of foreign accent does not always match with ratings of listener effort required in understanding the speech [3]. For example, in Derwing and Munro [3], native English listeners were presented with spoken sentences produced by foreign-accented talkers from four language background (Cantonese, Japanese, Polish, and Spanish) and were asked first to transcribe the utterances, then to rate their degree of foreign accent (from “no accent” to “extremely strong accent”) and after a short break to rate how easily they could understand them (from “extremely easy to understand” to “extremely difficult or impossible to understand”). The results showed that the accent ratings tended to be more extreme than the ease of understanding ratings, which were harsher than intelligibility scores (based on transcription errors). This indicates that the salience of an accent in foreign speech may not be commensurate with the effort required to understand it or its intelligibility.

In evaluating how foreign accent affects speech perception (ease and intelligibility) it is important to consider both the

segmental and suprasegmental levels. Even though previous results show that the intelligibility of L2 speech and the extent that a foreign accent is perceived can be explained by differences at both segmental [e.g., 4-7] and prosodic levels [e.g., 8-10], the bulk of research has focused on the possible effects from segmental differences. For example, extensive studies have shown the difficulties in producing non-native phonemes such as English /l-r/ distinction by Japanese learners of English [e.g., 4]. Research in L2 vowel production also showed the influence of different vowel inventory size between L1 and L2 [6]. L2 learners face difficulty in producing non-native vowel contrasts especially among the learners with a smaller vowel inventory (e.g., five in Japanese) when producing vowels with a larger vowel inventory (e.g., Australian English) [7].

Relatively fewer studies have examined how non-native suprasegmental differences contribute to foreign accent. Such difference can be glossed in terms of L2 speakers having a non-native prosody. Speech prosody refers to features of speech such as pitch, stress, and duration [11], and these features also differ cross-linguistically. For example, differences in how the rhythms of languages pattern, have led to the idea of different rhythm classes. This typology groups languages into so called stress-timed (e.g., English, Dutch), syllable-timed (e.g., Spanish, French) and mora-timed (e.g., Japanese) classes [12, 13].

By and large, it has been found that the characteristics of L2 rhythm are influenced by the L1 rhythm type [e.g., 8, 9]. These studies have used duration-based measures to characterize L2 rhythm, however, recent studies suggest that patterns in speech amplitude also provide an index of rhythm. For example, a rhythm metric has recently been developed on the basis of changes in speech amplitude. This metric uses the amplitude modulation (AM) structure of the speech envelope and was derived from the Spectral Amplitude Modulation Phase Hierarchy (S-AMPH) model [14]. This rhythm metric is different from previous durational statistic-based approaches since it only considers slow changes (modulations) in the amplitude of acoustic signals. Our preliminary analysis using this model suggests differences in AM between L1 English and L2 English (produced by Japanese talkers). While further studies are required, different amplitude patterns between L1 and L2 talkers likely exist and so will contribute to foreign-accent. Taking the above into account, in the current study we investigated the extent to which L2 speech rhythm affects the ease of perceiving speech and the degree of foreign accent and we evaluated L2 rhythm using both duration and amplitude measures.

Before detailing the experimental method, we consider one other prominent factor that has been found to affect the perceived degree of foreign accent, L2 talker’s experience. Models of learning L2 speech (e.g., Speech Learning Model (SLM) [15, 16] and Perceptual Assimilation Model of Second

Language Speech Learning (PAM-L2) [17]) have postulated that L2 learners will improve in their perception of non-native contrasts with experience of the L2 speaking environment. Consistent with this, research comparing Japanese learners of L2 (English) vowels with more than 12 months of residence in an English speaking country to those with 6 months or less, has shown significant improvement in the former compared to the latter [7] (although this was not the case for difficult contrasts, e.g., tense-lax vowels). Given these models and studies have focused on segmental learning, the extent to which L2 prosody can be acquired is still largely unknown. Thus, it is important to investigate how L2 prosody is produced and perceived as a function of L2 experience.

One study has investigated the influence of L2 experience on speech intelligibility and accentedness with English-French bilinguals [18]. In this study, the speech of L2 talkers that had different experience levels (inexperienced and experienced French talkers of English) was used as the basis for a prosody ‘transplantation’ study. In this approach, certain prosodic features (segment duration, amplitude and F0 contour) from the L2 English utterances were transplanted onto ones spoken by native English speakers. Then these sentences were tested for intelligibility and rated for foreign accent. The results showed that transplanting features of the inexperienced French speaker’s English prosody onto English native segments decreased intelligibility and increased accentedness ratings.

The current study followed up [18] by investigating the effect of non-native speech rhythm on the perceived degree of foreign accent with Japanese-accented English as a function of L2 experience. In order to assess the influence L2 rhythm on the perception of foreign-accented speech, we used a Pitch-synchronous overlap-add (PSOLA) procedure [21] that allowed the L2 duration and amplitude produced of an experienced and an inexperienced Japanese talker to be transplanted onto the segments produced by a native Australian English talker. The transplanted speech was then presented to native Australian English listeners who were asked to rate the degree of foreign accent (accent rating task) and how easily they could understand (ease of understanding; also known as comprehensibility [1-3]) that speech using a 9-point scale. It was expected that perceived degree of foreign accent and comprehensibility would be affected by L2 talker’s experience. Specifically, higher accentedness and reduced ease of understanding were predicted for the inexperienced Japanese-English L2 speech compared to the more experienced ones.

## 2. Methods

### 2.1. Participants

Forty five native Australian English listeners (34 females, 11 males;  $M_{age} = 22.5$  years) participated in this study. They were recruited from the Western Sydney University using the university’s research participation system. All of the participants reported normal hearing. From a questionnaire we ascertained that none of the participants were familiar with Japanese-accented English.

### 2.2. Stimuli

#### 2.2.1. Materials/Talkers

The stimuli consisted of 56 IEEE Harvard Sentences produced by two Japanese talkers and one Australian English talker (all females;  $M_{age} = 24.0$  years) who resided in Sydney, Australia.

The Japanese talker who was considered to be “inexperienced” in English had fairly recently arrived in Sydney (length of residence (LOR) = 4.5 months) and the “experienced” Japanese talker had been in Sydney more than a year (LOR = 12.5 months). According to our separate foreign accent rating study by native Australian English listeners ( $n = 15$ ), the inexperienced Japanese talker’s English was perceived as “strongly foreign-accented” (8.1 out of 9) and the experienced Japanese talker’s English was perceived as “mildly foreign-accented” (4.7 out of 9). The monolingual Australian English talker was born and raised in Sydney, and was recruited at Western Sydney University. All talkers reported no history of speech, vision or hearing problems.

#### 2.2.2. Stimulus editing (PSOLA)

The recording was made using an externally connected lapel microphone, (an AT4033a audio-technica microphone) in 44.1 kHz, 16-bit mono. Following the recording, all the stimulus sentences were segmented into phonemes. A Pitch-synchronous overlap-add (PSOLA) procedure [21] was then used to impose L2 durations and amplitude onto the native segments. Table 1 describes the details of how segments and prosody were combined: (1) native English segments with inexperienced Japanese amplitude (NE\_NJI:I); (2) native English segments with inexperienced Japanese durations (NE\_NJI:D); (3) native English segments with experienced Japanese amplitude (NE\_NJE:I); and (4) native English segments with experienced Japanese durations (NE\_NJE:D). In addition to the four manipulated conditions, another condition, native English (no manipulation), was also prepared.

Table 1. *Illustration of stimulus manipulations*

Segment	Transferred Prosody	Manipulation Types	Codes
Native English	NA	NA	NE
	Inexperienced Japanese	Amplitude	NE_NJI:I
		Duration	NE_NJI:D
	Experienced Japanese	Amplitude	NE_NJE:I
Duration		NE_NJE:D	

### 2.3. Procedure

#### 2.3.1. Accent Rating Task

The participants were tested individually in a sound-treated booth. All of the participants performed an accent rating task [cf. 1-3]. They were required to listen to the English sentences while paying careful attention to the accent. They were then asked to rate the degree of accent on a 9-point scale (from 1: no foreign accent at all to 9: very strong foreign accent). They were encouraged to use the full rating scales for the judgments. Both rating tasks started with practice trials ( $n = 4$ ) to become familiar with the rating scales.

#### 2.3.2. Comprehensibility Rating Task

After completing the accent rating task, a comprehensibility (ease of understanding) rating task was conducted. In this task, the participants were instructed to listen to each sentence and were asked to rate how easily they could understand it using a 9-point scale (from 1: easy to understand to 9: very difficult to

understand). Here, the intention was to measure listener effort in the perception of foreign-accented speech [cf. 2]. Note, the same stimulus materials were used as the accent rating task.

### 3. Results

Separate analyses for accent rating and comprehensibility rating data were conducted with one-way ANOVAs with `lm` in R 3.2.1. to predict accent rating score (3.1) and comprehensibility rating score (3.2) based on different prosodic manipulation types.

#### 3.1. Accent Rating

Figure 1 shows mean accent rating scores for each prosody manipulation condition. Across the participants, the mean accent rating score in Native English (both segments and prosody) was perceived lower (less accented) than the rest of manipulated conditions, NE\_NJE:D ( $\beta = 0.75$ , SE = 0.08,  $t = 9.08$ ,  $p < .0001$ ), NE\_NJI:D ( $\beta = 0.99$ , SE = 0.08,  $t = 11.962$ ,  $p < .0001$ ), NE\_NJE:I ( $\beta = 0.30$ , SE = 0.08,  $t = 3.59$ ,  $p < .0001$ ), NE\_NJI:D ( $\beta = 0.26$ , SE = 0.08,  $t = 3.116$ ,  $p < .001$ ). Furthermore, we ran multiple comparisons between stimulus manipulation types (duration and amplitude) and L2 experience types (experienced vs. inexperienced) using the `glht` function in R package `multcomp` [22].

The results showed stronger foreign-accented durational influences compared to the amplitude ones on perceived degree of foreign accent in both experienced Japanese ( $\beta = -0.46$ , SE = 0.10,  $t = -4.76$ ,  $p < .0001$ ) and inexperienced Japanese ( $\beta = -0.73$ , SE = 0.10,  $t = -7.76$ ,  $p < .0001$ ). In addition, we found that the experienced Japanese non-native prosody was perceived less accented compared to the inexperienced Japanese only in the duration manipulation ( $\beta = 0.24$ , SE = 0.10,  $t = 2.50$ ,  $p < .05$ ), but not in the amplitude one ( $p > .05$ ).

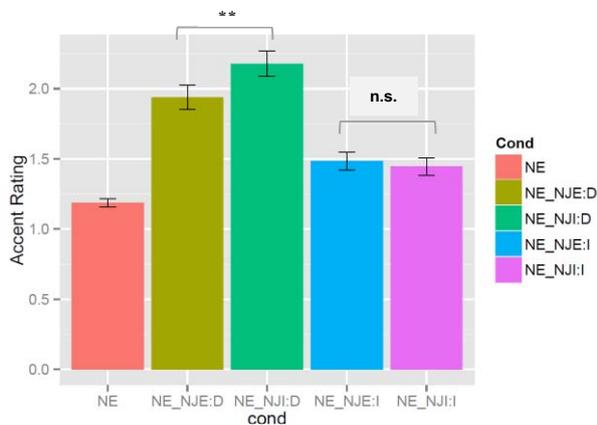


Figure 1: Mean accent rating scores for each prosody manipulation conditions. (NE: Native English, NE\_NJE:D: Native English segments with experienced Japanese durations, NE\_NJI:D: Native English segments with inexperienced Japanese durations, NE\_NJE:I: Native English segments with experienced Japanese intensity, and NE\_NJI:I: Native English segments with inexperienced Japanese intensity). Error bars indicate +/- one standard error.

#### 3.2. Comprehensibility Rating

Figure 2 shows mean comprehensibility rating scores for each prosody manipulation condition. The mean comprehensibility

rating score in Native English (both segments and prosody) was lower (i.e., easier to be understood) compared to the other manipulated conditions, NE\_NJE:D ( $\beta = 0.67$ , SE = 0.76,  $t = 8.78$ ,  $p < .0001$ ), NE\_NJI:D ( $\beta = 0.84$ , SE = 0.76,  $t = 10.97$ ,  $p < .0001$ ), NE\_NJE:I ( $\beta = 0.35$ , SE = 0.76,  $t = 4.56$ ,  $p < .0001$ ), NE\_NJI:D ( $\beta = 0.31$ , SE = 0.08,  $t = 4.11$ ,  $p < .001$ ). As with the previous analyses, we ran multiple comparisons between stimulus manipulation types (duration and amplitude) and L2 experience types (experienced vs. inexperienced).

Similar to the accent rating data, there were stronger durational influences compared to the amplitude ones on comprehensibility in both experienced Japanese ( $\beta = -0.32$ , SE = 0.08,  $t = -3.6$ ,  $p < .0001$ ) and inexperienced Japanese ( $\beta = -0.52$ , SE = 0.08,  $t = -5.93$ ,  $p < .0001$ ). However, unlike the accent rating scores, we found no significant difference in comprehensibility scores between the experienced and the inexperienced Japanese stimuli in both duration and amplitude ( $p > .05$ ).

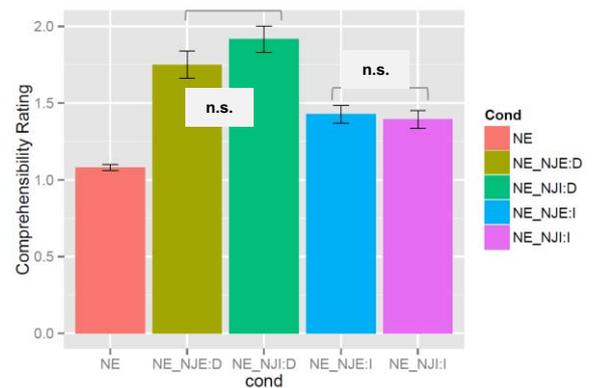


Figure 2: Mean comprehensibility rating scores for each prosody manipulation conditions. (NE: Native English, NE\_NJE:D: Native English segments with experienced Japanese durations, NE\_NJI:D: Native English segments with inexperienced Japanese durations, NE\_NJE:I: Native English segments with experienced Japanese intensity, and NE\_NJI:I: Native English segments with inexperienced Japanese intensity). Error bars indicate +/- one standard error.

### 4. Discussion and conclusions

The aim of this study was to investigate the effect of non-native speech rhythm in the perception of foreign accentedness and comprehensibility. Speech rhythm is typically characterized using duration-based metrics [e.g., 8, 9], but a recent study suggests an important role of amplitude as well [14]. We investigated the influences of both of these properties by transplanting L2 (Japanese speaking English) speech timing and amplitude values onto native (English) segment productions. We also examined how different degrees of L2 English experience (i.e., experienced vs. inexperienced) among the Japanese talkers affected the perceived degree of foreign accent and comprehensibility (i.e., ease of understanding).

The results showed that non-native speech duration affected both accentedness and ease of understanding. That is, the native English listeners perceived L2 duration transplanted sentences as being more foreign-accented and more difficult to understand than untransplanted (control) sentences. The talker's L2 experience had an effect only on the perceived

degree of foreign accent, but not on listening effort. This result is consistent with previous research showing that listeners are more sensitive to foreign accent than the effort to understand [3]. We suggest that the less experienced L2 talker's sentences were rated having more foreign accent due to their timing being more variable. In our previous acoustic analyses, we found that L2 vowel duration variability (rhythm) was larger for the inexperienced compare to the experienced L2 talkers [19]. Thus, the increase in perceived accentedness for sentences transplanted with the inexperienced Japanese talker's duration may be due to native listener's sensitivity to changes in durational patterns.

Transplanting non-native speech intensity did not produce the same effect on perception as transplanting duration. For example, although L2 talker's experience made a difference for the perception of duration-transplanted speech, transplanted intensity did not. This difference between timing and intensity may involve the degree to which these factors are apparent in acoustic differences between the two talkers and whether these are perceptually salient. That is, there may be little difference in intensity between the experienced and inexperienced L2 talkers. To our knowledge, there is no study that has examined L2 intensity as a function of L2 experience, and our future analyses will address this. In addition, intensity changes may be less perceptually salient than durational ones. Indeed, the effect of non-native intensity was smaller compared to the duration effect for both accentedness and listener effort, and much smaller compared to segmental contributions found in previous research [23, 24]. It is of course possible that although the behavioural effect is subtle, there could be a large effect in brain-based measures. That is, less efficient neural entrainment may occur with non-native intensity patterns considering recent neurophysiological findings [e.g., 25]. Further research is necessary to address this issue.

Overall, our findings highlight the role of non-native duration patterns on the perceived degree of foreign accent as a function of L2 experience. We are currently conducting experiments examining intensity measures for the English production by the experienced and inexperienced L2 talkers, aiming to understand how L2 speech rhythm is acoustically different from native speech rhythm and how the difference affects perception. We believe that the current results offer an insight into the many ways that L2 rhythm can affect the perception of foreign-accent.

## 5. References

- [1] Munro, M., and Derwing, T., "Foreign accent, comprehensibility, and intelligibility in the speech of second language learners", *Language Learning*, 45, 73-97, 1995.
- [2] Derwing, T. M., and Munro, M.J., "Putting accent in its place: Rethinking obstacles to communication", *Language Teaching* 42, 276-490, 2009.
- [3] Derwing, T. M., and Munro, M. J., "Accent, intelligibility, and comprehensibility: Evidence from four L1s", *Studies in Second Language Acquisition*, 19, 1-16, 1997.
- [4] Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., and Fujimura, O., "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," *Perception & Psychophysics*, vol. 18, 331-340, 1975.
- [5] A. Sheldon, and W. Strange., "The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception," *Applied Psycholinguistics*, vol. 3(3), 243-261, 1982.
- [6] Iverson, P., and Evans, B. G., "Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration" *The Journal of the Acoustical Society of America*, 122(5), 2842-2854, 2007.
- [7] Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., and Yamada, T., "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), 156-157, 2011.
- [8] White, L., and Mattys, S. L., "Calibrating rhythm: First language and second language studies," *Journal of Phonetics*, vol. 35, (4), pp. 501-522, 2007.
- [9] Ordin, M., and Polyanskaya, L., "Development of timing patterns in first and second languages," *System*, 42, 244-257, 2014.
- [10] Aoyama, K., and Guion, S. G., "Prosody in second language acquisition: Acoustic Analyses of duration and F0 range". In Bohn, O.-S., Munro, M.J. (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, Amsterdam, John Benjamins, 282-297, 2007.
- [11] Lehiste, I., "Suprasegmentals", Cambridge, MA: MIT Press, 1970.
- [12] Abercrombie, D., "Elements of general phonetics" (Vol. 203): Edinburgh University Press Edinburgh, 1967.
- [13] Pike, K. L., "The Intonation of American English", 1945.
- [14] Leong, V., "Prosodic rhythm in the speech amplitude envelope: Amplitude modulation phase hierarchies (AMPHs) and AMPH models", Cambridge, UK: PhD Thesis, 2012.
- [15] Flege, J. E., "Second-Language Speech Learning: Theory, Findings and Problems", in W. Strange [Ed], *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, York, Timonium, MD, 233-273, 1995.
- [16] Flege, J. E., "Interactions between the native and second-language phonetic systems", *An Integrated View of Language Development: Papers in Honor of Henning Wode*, Trier Wissenschaftlicher Verlag., 217-244, 2002.
- [17] Best, C.T. and Tyler, M.D., Nonnative and second-language speech perception: Commonalities and complementarities, in *Second language speech learning: The role of language experience in speech perception and production*, M.J. Munro and O.-S. Bohn, Eds. 2007, John Benjamins: Amsterdam. 13-34.
- [18] Pinet, M., and Iverson, P., "Talker-listener accent interactions in speech-in-noise recognition: Effects of prosodic manipulation as a function of language experience", *The Journal of the Acoustical Society of America*, 128(3), 1357-1365, 2010.
- [19] Kawase, S., Kim, J., and Davis, C., "The influence of second language experience on Japanese-accented English rhythm," *Proceeding of Speech Prosody 2016*, Boston, USA.
- [20] Mochizuki-Sudo, and S. Kiritani., "Production and perception of stress-related durational patterns in Japanese learners of English," *Journal of Phonetics*, vol. 19(2), 231-248, 1991.
- [21] Yoon, K., "Imposing native speakers' prosody on non-native speakers' utterances: The technique of cloning prosody", *Journal of the Modern British & American Language & Literature*, 25(4), 197-215, 2007.
- [22] Hothorn, T., Bretz, F., and Westfall, P., "Simultaneous inference in general parametric models", *Biometrical journal*, 50(3), 346-363, 2016.
- [23] Winters, S, and O'Brien, M. G., "Perceived accentedness and intelligibility: The relative contributions of F0 and duration", *Speech Communication*, 55 (3), 486-507, 2013.
- [24] Rognoni, L., and Busà, M. G., "Testing the Effects of Segmental and Suprasegmental Phonetic Cues in Foreign Accent Rating: An Experiment Using Prosody Transplantation", *Proceedings of the International Symposium on the Acquisition of Second Language Speech Concordia Working Papers in Applied Linguistics*, 5, 547-560, 2014
- [25] Schroeder, C. E., and Lakatos, P., "Low-frequency neuronal oscillations as instruments of sensory selection", *Trends in Neurosciences*, 32(1), 9-18, 2009.