THE PERCEPTION OF JAPANESE LEXICAL CONTRASTS BY L2 LEARNERS: THE ROLE OF THE LEXICON

Marco Fonseca

University of Illinois at Urbana-Champaign
marcofon@illinois.edu

ABSTRACT

This paper examines the perception of three types of Japanese lexical contrasts by L2 learners native speakers of English: words with pitch accent on the penultimate syllable versus words with pitch with on the last syllable, voiceless stops versus voiced stops, and phonemically long versus short vowels. An ABX discrimination task and a lexical assignment task were conducted. A group of 22 L2 learners and a control group of 17 native speakers of Tokyo Japanese were recruited. The results of this study show that although L2 learners are able to discriminate the three types of contrast as well as native speakers, they find it difficult to assign meaning to different contrasts, especially for pitch accent. It is argued that this is evidence that the lexical pitch accent contrast is "weakly" encoded by L2 learners.

Keywords: speech perception, lexical contrast, second language acquisition, pitch accent, Japanese.

1. INTRODUCTION

This paper examines how second language learners native speakers of English (henceforth L2 learners) perceive three different types of Japanese lexical contrasts: voiceless versus voiced stops: *kakkou* "outfit" versus *gakkou* "school"; phonemically long versus short vowels: *biru* "building" versus *biiru* "beer"; and words with lexical pitch accent on their last syllable versus first syllable: *háshi* "chopsticks" versus *hashí* "bridge". The following research questions are proposed: are pitch suprasegmental contrasts more difficult to be perceived than segmental contrasts (RQ1)? Can L2 learners encode (i.e., assign meaning to) these different types of lexical contrast (RQ2)? The motivation for choosing English and Japanese and the importance of these two research questions are as follows. The phonetic implementation of these three type of contrasts in Japanese is different from English: English uses a lax-short vowel distinction [1], is a stressed-timed language [2], and has longer voice onset time than Japanese [3]. Moreover, previous studies have generally focused on a specific type of contrast, such as vowel length [4, 5], voicing of stops [6, 7], or pitch accent [8, 9]. In this study, three different types of contrasts were analyzed. This allows us to determine the relative difficulty in the acquisition of prosodic contrasts (e.g., pitch accents) with respect to segmental contrasts (voicing of stops and vowel length). Furthermore, previous studies have focused on the perception of phonetic differences and on the influence of one's native language phonological categories on the perception of L2 categories [10, 11, 12]. Fewer studies have paid attention to whether L2 learners can encode contrastive L2 categories in their mental lexicon [13, 14, 15]. Therefore, the importance of the research questions is two-fold: it will be tested whether different contrasts are perceived differently by L2 learners, since previous studies tend to focus solely on one type of contrast. It will also be tested the role of lexical encoding (i.e., whether L2 learners are able to assign meaning to different minimal pairs of different contrast types or not).

If the answer to RQ1 is no and even so L2 learners have a hard time encoding different lexical contrasts (RQ2), perceiving an L2 contrast does not imply that the contrast will be learned, as evidenced by Amengual [13] and Ota et. al [14]. This would be evidence that the phonology of an L2 is acquired differently from that of an L1, since native speakers do not seem to find difficulty in encoding different lexical items. To answer these two research questions an ABX discrimination task and a lexical assignment task were conducted. If pitch accent contrasts are more difficult to be acquired than segmental contrasts, participants will perform better with stimuli containing minimal pairs differing in pitch accent than the other contrasts. Furthermore, if L2 learners find it more difficult to encode pitch accent in comparison to stop and vowel contrasts, they will find it more difficult to assign meaning to minimal pairs differing in pitch accent than minimal pairs differing in vowel length and stop voicing.

2. METHODS

Thirty-nine participants were recruited. From these, 22 were advanced L2 learners native speakers of English (7 females and 15 males). Seven were recruited at the University of Illinois and 15 were recruited remotely and performed the experiment online. From these 15, 13 were currently residing in the Tokyo region and 2 had recently returned to the U.S. from a year abroad studying in the Tokyo region. All participants residing in the U.S. had spent at least one year studying in Japan at the university level. The range of all L2 learners' time studying Japanese was 2-18 years (M: 6.42, SD: 4.75). Their age range was 21-42 years old (M: 27.95, SD: 6.7). The age that they started studying Japanese ranged from 13-25 years old (M: 18.77, SD: 3.22). For the participants living in Tokyo, their age of arrival ranged from 20 to 31 years (M: 24, SD: 3.015). Eighteen participants were from the U.S., 3 were from the U.K. and 1 was from Australia. All of them self-rated as being advanced learners[1]. The control group was comprised of 17 native speakers of Tokyo Japanese (5 females and 12 males). Six of them were recruited at a research university in the U.S. and the remaining 12 were recruited remotely and were living in the Tokyo region. All native speakers were university students. Their age range was 21-40 (M: 28.94, SD: 5.56). None of the participants reported having been diagnosed with a hearing impairment.

A total of 60 words, comprised of 10 minimal pairs for each type of contrast (pitch accent, stops, and vowel length) were selected. Pitch accent words minimal pairs were accented in their first syllable. Some words were unaccented rather than second accented. However, because of the initial lowering rule in Tokyo Japanese, and because they were phrase-final, without any following particles, their phonetic realization was the same as second accented words. Eight of the minimal pairs were disyllabic. The remaining two were trisyllabic. Each minimal pair of voiced/voiceless stops had the same accent pattern. For the stop contrasts, only minimal pairs of voiceless/voiced velars /k, g/ were selected. Nine of the minimal pairs had the velar stop in the onset of the first syllable. The remaining one had the velar stop in the onset of the second syllable. Words of a previous study by Hirata [4] were selected for long and short vowel minimal pairs.

The minimal pairs were recorded by two native speakers of Tokyo Japanese who resided in the U.S. at the time of the recording. One participant self-identified as a female and the other one as a male. Native speakers of different genders were recorded to increase the difficulty of the task: in the ABX, A and B were produced by the male participant and X was produced by the female participant. The female participant was in her early thirties at the time of the recording while the male participant was in his late twenties. The native speakers were recorded at a sound-attenuated booth. An AKG C520 head-mounted microphone positioned approximately 2 inches away from the participants' mouth was used. The audio was captured by a Marantz PMD570 solid state recorder with a sampling rate of 48 kHz and a sample size of 16 bits. Participants were asked to produce each word 5 times and the researcher selected the token that sounded most representative of the category.

An ABX discrimination task and a lexical assignment task were conducted. For the participants recruited at the research university, the software PsychoPy2 (version 1.85.4) and Sony Dynamic Stereo Headphones MDR-7508 were used. For the participants recruited online, the online platform Qualtrics was employed. It was asked for the participants to conduct the experiment with ear/headphones. In the ABX, participants heard the first word ("A"), the second word ("B"), and then the third word ("X"). Then the following sentence would appear on the screen: "Is the last word you heard the same as the first or as the second?". If the third word was the same as the first one, participants would press the left arrow key. If the third was the same as the second one, they would press the right arrow key. Before starting the task, a quick training was performed in order to get the participants familiar with the experiment. After the training session, they heard the 30 minimal pairs in 4 word orders (ABA, ABB, BAA, and BAB), for a total of 120 items. There was an interval of 2 seconds between each word. In the lexical assignment task, participants heard the same words that they heard in the ABX task (produced by the male speaker), one at time. Along with the audio, (e.g. */háçi/*), they were presented two options (e.g. *chopsticks* or *bridge*), and were asked to select the appropriate meaning of the word[2]. The words that they reported not knowing were deleted from the analysis. Each participant took approximately 60 minutes to complete the tasks.

3. RESULTS

A total of 4560 responses were analyzed in the ABX task. From these, 2040 (120 tokens * 17 participants) were obtained from native speakers

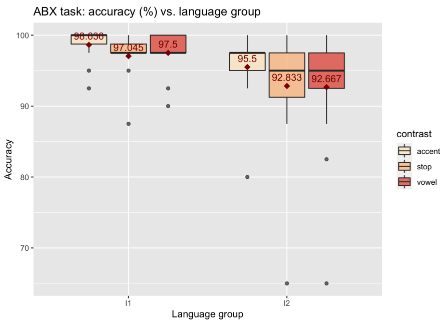

Figure 1: Accuracy by language group and type of contrast for the ABX task. Red numbers in the boxes are the means for each group.

while 2520 (120 tokens * 21 participants) were obtained from L2 learners. For each type of contrast, 840 (2520 responses /3 types of contrasts) correct responses were possible for the L1 group and 280 (840 responses /3 types of contrasts) correct responses were possible for the L2 group. The results of the ABX task are illustrated in Figure 1.

Both language groups performed at ceiling in the three types of contrast tested. The data were fit into a linear mixed effects logistic regression (function *glmer*, package *lme4* [16]) in R [17]. P-values were obtained with the function *mixed* from the package *afex* [18]. The dependent variable was response (two levels: wrong/right); the predictors were language group (two levels: L1/L2), type of contrast (three levels: accent/stops/ vowels) and their interaction, experimental environment (two levels: in person/remote); random effects were participant and item. All of the mixed effects regressions presented in this paper had their random effects maximized [19]. Considering a statistical significance threshold of $p < 0.05$, no significance was found for any of the predictors [3].

For the lexical assignment task, 2340 responses were obtained (60 items * 39 participants). From these responses, 1020 (60 items * 17 participants) were from native speakers and 1320 were from L2 learners (60 items * 7 participants). For each type of contrast, 340 (1020 responses / 3 types of contrasts) correct responses were possible for the L1 group and 440 (1320 responses / 3 types of contrasts) correct responses were possible for the L2 group. The distribution of correct responses is illustrated in Figure 1. In this task, the L1 group performed at ceiling. For L2 learners, on the other hand, a greater discrepancy was found. The lowest accuracy was observed for accent contrast (70%), followed by stops (86%), and vowels (87%). Similarly to the ABX, the responses were fit into a linear mixed effects logistic regression. The dependent variable was

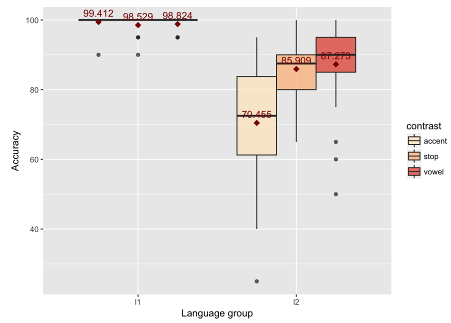

Figure 2: Accuracy by language group and type of contrast for the lexical assignment task.

response (two levels: wrong/right); the predictors were language group (two levels: L1/L2), type of contrast (three levels: accent/stops/vowels) and their interaction, experimental environment (two levels: in person/remote); random effects were participant and item. There was an effect of language ($\chi^2(1)$ = 71.12, $p < 0.0001$) and its interaction with contrast ($\chi^2(2) = 8.06$, $p = 0.02$). Post-hoc comparisons were conducted using the *emmeans* function from the package *emmeans*. Almost all of the comparisons (p-values ranging from < 0.0001 to 0.0041) were significant. All estimates of all contrasts for the native speakers were higher than the estimates of L2 learners. This indicates that native speakers performed better than L2 learners across the three contrasts tested in this experiment. For L2 learners the comparison between the accent contrast versus the vowel contrast was significant, and the estimate of the former contrast (the baseline) was lower than the latter contrast (β = -1.23, p = 0.004). This indicates that the vowel contrast was easier to be discriminated than pitch accent. The same tendency was observed in the pitch accent versus stop contrast for L2 learners (β = -1.13, p = 0.01). Thus, the statistics indicate that the accent contrast is the most difficult one to be identified in the lexical assignment task for L2 learners.

4. DISCUSSION

The first RQ proposed by this paper is whether suprasegmental contrasts are more difficult to be perceived than segmental contrasts by L2 learners. The results of the ABX task showed that L2 learners are able to perceive the difference between segmental vs. suprasegmental contrasts as well as native speakers. The second RQ is whether L2 learners encode these different types of lexical contrast. The results of the lexical assignment task showed that although L2 learners can accurately assign meaning to minimal pairs differing in stop and vowel contrasts as well as native speakers, they find it difficult to assign meaning to pitch accent minimal pairs. The results of the present paper have implications for models on the acquisition of L2 phonology. This section discusses some of these models, and whether their predictions can explain the results of the present paper.

Both the Native Language Magnet Model (NLM, [20]) and the Speech Learning Model (SLM, [21, 11]) take into account the similarity of phonetic properties of an L1 sound and an L2 sound. According to NLM, an L1 sound whose phonetic properties are similar to an L2 sound would work as a prototype. This prototype would act as a perceptual magnet that attracts L2 sounds, which would be categorized based on this prototype. As L1 sound categories develop through adulthood, the model proposes that these categories will become more powerful attractors of L2 categories. According to SLM, the sounds that comprise an L1 and L2 phonetic system of a bilingual share a common phonological space. Consequently, these two systems will influence one another. SLM proposes that speech learning abilities remain intact throughout a learner's lifespan. However, age effects arise due to how the L1 and L2 phonetic systems interact. A revised version of SLM (SLM-r, [22]) switches the focus from age effects to the precision of the L1 category: through time an L1 category becomes more robust and therefore it will more likely affect an L2 category. These two models take into account phonetic similarities between two languages, however, encoding of lexical items is outside their scope and therefore they cannot fully explain the results of the present paper.

Another prominent model regarding the perception of nonnative speech categories is the Perceptual Assimilation Model—PAM-L2 [23, 24, 25]. Similarly to NLM and SLM, PAM-L2 argues that not all sound categories of a nonnative language are acquired equally. However, a major difference between PAM-L2 and the other models is that PAM-L2 proposes that listeners are able to detect articulatory gestures. Within articulatory phonology [26], gestures are considered the atomic unit of representation. Thus, while SLM and NLM focus on the phonetic similarities between L1 and L2 sounds, PAM-L2 taps into the phonological level. Although PAM-L2 does not explicitly discuss lexical encoding, positing the importance of the phonological level provides it a level of abstraction.

The Second Language Linguistic Perception Model—L2LP [27, 28] also takes into account the learner's native phonological system. This model posits that L2 learners copy their native

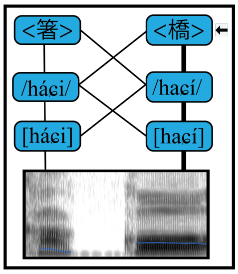

Figure 3: Levels of L2 sound categories acquisition as proposed by the revised L2LP. This diagram illustrates the correct acquisition of 橋 /haɕí/ "bridge", rather than 箸 /háɕi/ "chopsticks".

phonological system and use this copy in order to develop new sound categories. From there, listeners might create a new sound category for the L2 sound if it is considered to be too different from L1 categories. It might also be the case that a category present in the L1 phonological system is reused in order to categorize the L2 sound. A revised version of this model further proposes that learning an L2 sound is divided into different levels. According to the revised L2LP, learning of an L2 phonetic category is meaning-driven and occurs sequentially: acoustic form → [phonetic form] → /phonological form/ → <lexical form>. Such levels are connected and input strengthens these connections, as shown in Figure 3. It is argued that this model can best explain the results of the present paper.

The results of the perception experiment show that as argued by these models, the perception of lexical contrasts varies between native speakers and L2 learners, and that such differences might be because of phonetic differences. Furthermore, meaning also plays a role in the accuracy of this perception, highlighting the importance of the meaning-driven learning of L2 phoneme categories.

5. CONCLUSION

The primary finding of this study is that even if L2 learners have learned to attend to L2 contrasts that have a phonetic implementation different from their L1, they may still fail to fully use that distinction in their mental lexicon. The present study provided evidence that even if L2 learners can hear the difference between a lexical contrast that are not present in their native language, it does not mean that their lexical representation is as robust as the lexical items of their native language. Such differences could be because pitch accent differences are not commonly taught in the classroom, or because Japanese orthography does not indicate it. These may be worthwhile factors to explore in future research.

REFERENCES

[1] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *The Journal of the Acoustical Society of America*, vol. 24, no. 2, pp. 175–184, 1952.

[2] M. E. Beckman and J. Edwards, "Articulatory evidence for differentiating stress categories," *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form*, pp. 7–33, 1994.

[3] T. J. Riney, N. Takagi, K. Ota, and Y. Uchida, "The intermediate degree of VOT in japanese initial voiceless stops," *Journal of Phonetics*, vol. 35, no. 3, pp. 439–443, 2007.

[4] Y. Hirata, "Effects of speaking rate on the vowel length distinction in Japanese," *Journal of Phonetics*, vol. 32, no. 4, pp. 565–589, 2004.

[5] M. Nakagawa and T. Funamura, "Shokyuu nihongo gakushuusha no choutan boin no ninshiki keikou to jizoku jikan [auditory perception of long and short vowels by beginning learners of Japanese in relation to duration time]," *Journal of International Student Center, Hokkaido University*, vol. 4, pp. 18–37, 2000.

[6] N. Yamada, "Nihongo akusento shuutoku no ichi dan kai - gaikokujin gakushuusha no baai [The steps of acquisition of Japanese accent - the case of foreign learners]," *Nihongo Kyouiku*, vol. 83, pp. 108–120, 1994.

[7] M. Fukuoka, "Pekingo bogo washa to shanhaigo bogo washa o taishou to shita nihongo no yuusuei musei haretsuon no oudanteki shuutoku kenkyuu [a longitudinal study in the acquisition of voiced and voiceless stops by native speakers of Beijing Mandarin and Shanghainese]," *Nihongo Kyouku*, no. 87, pp. 40–53, 1995.

[8] T. Ayusawa, "Gaikokujin gakushuusha no nihongo akusento intoneeshon shuutoku (acquisition of Japanese accent and intonation by foreign learners)," *Journal of the Phonetic Society of Japan*, vol. 7, no. 2, pp. 47–58, 2003.

[9] J. J. Atria and V. L. Hazan, "Development of accentual categories in japanese as a second language," *Proceedings of the 18th International Congress of Phonetic Sciences*, 2015.

[10] J. E. Flege, "The production of new and similar phones in a foreign language: Evidence for the effect of equivalence classification," *Journal of phonetics*, vol. 15, no. 1, pp. 47–65, 1987.

[11] ——, "Language contact in bilingualism: Phonetic system interactions," *Laboratory Phonology*, vol. 9, pp. 353–382, 2007.

[12] C. T. Best and M. D. Tyler, "Nonnative and second-language speech perception: Commonalities and complementarities," *Language Experience in Second Language Speech Learning: In honor of James Emil Flege*, vol. 1334, 2007.

[13] M. Amengual, "The perception and production of language-specific mid-vowel contrasts: Shifting the focus to the bilingual individual in early language input conditions," *International Journal of Bilingualism*, vol. 20, no. 2, pp. 133–152, 2016.

[14] M. Ota, R. J. Hartsuiker, and S. L. Haywood, "The KEY to the ROCK: Near-homophony in nonnative visual word recognition," *Cognition*, vol. 111, no. 2, pp. 263–269, 2009.

[15] J.-W. Van Leussen and P. Escudero, "Learning to perceive and recognize a second language: The L2LP model revised," *Frontiers in psychology*, vol. 6, p. 1000, 2015.

[16] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[17] R. C. Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016. [Online]. Available: https://www.R-project.org/

[18] H. Singmann, B. Bolker, J. Westfall, and F. Aust, *afex: Analysis of Factorial Experiments*, 2016, r package version 0.16-1. [Online]. Available: https://CRAN.R-project.org/package=afex

[19] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of Memory and Language*, vol. 68, no. 3, pp. 255–278, 2013.

[20] P. K. Kuhl and P. Iverson, "Linguistic experience and the perceptual magnet effect," *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, pp. 121–154, 1995.

[21] J. E. Flege, C. Schirru, and I. R. MacKay, "Interaction between the native and second language phonetic subsystems," *Speech Communication*, vol. 40, no. 4, pp. 467–491, 2003.

[22] J. E. Flege and O.-S. Bohn, "The revised speech learning model (SLM-r)," *Second language speech learning: Theoretical and empirical progress*, pp. 3–83, 2021.

[23] C. T. Best, "The emergence of native-language phonological influences in infants: A perceptual assimilation model," *The development of speech perception: The transition from speech sounds to spoken words*, vol. 167, no. 224, pp. 233–277, 1994.

[24] ——, "A direct realist view of cross-language speech perception," *Speech perception and linguistic experience: Issues in cross-language speech research*, pp. 171–206, 1995.

[25] J. Chen, C. T. Best, M. Antoniou, and B. Kasisopa, "Cognitive factors in perception of Thai tones by naïve Mandarin listeners," *Proceedings of the 19th ICPHS, Melbourne*, pp. 1684–1688, 2019.

[26] C. P. Browman and L. Goldstein, "Articulatory phonology: An overview," *Phonetica*, vol. 49, no. 3-4, pp. 155–180, 1992.

[27] P. Escudero, "Linguistic perception of similar L2 sounds," *Phonology in Perception*, vol. 15, pp. 152–190, 2009.

[28] R. Mayr and P. Escudero, "Explaining individual variation in L2 perception: Rounded vowels in English learners of German," *Bilingualism: Language and Cognition*, vol. 13, no. 3, pp. 279–297, 2010.

[1] Logistically, it is very difficult to find L2 learners of Japanese who are native speakers of English, hence the somewhat heterogeneous L2 group. Several statistical analyses considering the possible differences in this group were conducted. No difference was found between such differences. These analyses are available in my GitHub account: https://github.com/marcomzp/ICPhSdata

[2] Participants were asked at the end of the experiment if there were words that they did not know. The words were chosen in consultation with instructors of Japanese who taught Japanese to some of the participants at the time

[3] All the data and analyses are also available in my GitHub account.