

Can acoustic cues used in L1 really be used to perceive novel sound contrasts?

Shiori Ikawa, Kumi Takimoto, Izabelle Grenon

The University of Tokyo

shiori.ikawa@gmail.com, kumichama9350@gmail.com, grenon@boz.c.u-tokyo.ac.jp

ABSTRACT

One issue in language acquisition is whether the acoustic cues used in the L1 can be used to perceive a different phonemic contrast in the L2. A previous study found that Japanese listeners can utilize their sensitivity to vowel duration in their L1 to identify the English coda voicing contrast when asked to classify each token as 'bit' or 'bid'. That is, they could use vowel duration in a way comparable to native English listeners when tested with a lexical identification task. The current study tested whether Japanese listeners could still utilize vowel duration to identify the same contrast when using a phoneme identification task: Japanese listeners were asked to categorize the last sound they heard as $\text{ɸ}ø$ or $\text{ɸ}dø$. Our results indicate that Japanese listeners rely on vowel duration only in the lexical identification task: the familiar acoustic cues are not applied to the novel phonemic contrast.

Keywords: second language, perception, levels of processing, cue weighting, identification task

1. INTRODUCTION

Recent models of second language (L2) sound perception posit different levels of speech processing (e.g., acoustic level and phonological level) that may play a role in the perception of L2 sound contrasts [6][9]. One question raised by multi-level models of L2 speech processing is whether acoustic contrasts can be dissociated from the phonological level to perceive novel L2 contrasts. For instance, Japanese speakers use vowel duration contrastively at the phonological level, where 'ie' *no* contrasts with 'ie' *home*. In [10] it was found that Japanese listeners could rely on vowel duration as an acoustic cue to contrast unfamiliar voiced-voiceless coda stop consonants in English in a way comparable to native North American English listeners when using a lexical identification task. The current study questions these results and its conclusion by conducting a similar experiment using a phoneme identification task.

In addition to the possible presence of periodicity during the word-final stop closure, native speakers of English use the duration of the preceding vowel as a cue to voicing: a short vowel indicates a

voiceless obstruent while a long vowel a voiced obstruent. Dutch, on the other hand, has an allophonic contrast of long and short vowels, where the high vowel is long before /r/ and short before other consonants [1]. However, as the voicing contrast is neutralized in word final position, this contrast is not used as a voicing cue for word final consonants [2]. A previous study examined the perception of the English coda voicing contrast by native Dutch listeners [2]. Stimuli were created from cross-splicing and editing naturally spoken nonce words, yielding 4 continua varying in the final coda consonant such as /s/ or /z/ and vowel duration (short to long). The participants were asked to judge what was the last sound they heard, $\text{ɸ}ø$ or $\text{ɸ}dø$ for instance. We will refer to this type of task as a *phoneme* identification task. The results indicated that Dutch listeners do not use vowel duration as a cue in a way comparable to English listeners.

Another study compared the perception of word-final voicing contrast by native listeners of English, Japanese, and Mandarin Chinese [3]. Japanese and Mandarin Chinese are similar in that they do not have word final stop consonants (with the exception of a nasal consonant in Japanese), but differ in that Japanese uses vowel duration phonemically as exemplified previously, while Mandarin does not. The stimuli in [3] were $\text{ɸ}otø$ and $\text{ɸ}odø$ in both natural production and in synthetic tokens manipulated in terms of F1 offset frequency and vowel duration. Participants were asked to judge whether the word they heard was $\text{ɸ}otø$ or $\text{ɸ}odø$ which we will refer to as a *lexical* identification task. The results show that while the F1 offset affected all listeners' perception, Japanese participants could use vowel duration as a cue to coda voicing contrast to a greater extent than Mandarin Chinese participants, although to a lesser extent than English participants.

Similar results were obtained when evaluating Japanese listeners' perception of the same English word-final voicing contrast [10]. This time the stimuli were created by manipulating a natural /bid/ sample in terms of vowel duration and duration of periodicity during the last stop closure, resulting in 24 $\text{ɸ}it$ - $\text{ɸ}idø$ stimuli. The participants completed a lexical identification task: they were asked to judge whether the word they heard was $\text{ɸ}itø$ or $\text{ɸ}idø$. The results indicated that Japanese listeners could use

vowel duration as well as the presence of periodicity to a similar extent than English listeners.

This brief review possibly suggests that acoustic cues processed at a phonemic level in one's L1 may be dissociated from the phonemic contrast in L1 and used to perceive novel contrasts—the Japanese case described above—as argued in [10]. Conversely, if the acoustic cues are processed only at a lower, phonetic level of processing, then this cue cannot be used to perceive novel L2 contrasts—the Dutch case cited above [2]. Another possibility, however, is that the differences in results are due to differences in experimental design. While all the studies above used an identification task, they differ in what is identified: phoneme [2] or word [3][10].

The purpose of this study was to see whether the type of task—phoneme vs lexical identification task—affects the perception of the word final voicing contrast in English by native Japanese listeners. Accordingly, we conducted a phoneme identification task with Japanese listeners, using the same 'bit-bid' stimuli as in [10], but asking the participants whether the last sound they heard was 't' or 'd'. Using the same design as in [10]—except for the choices given to the participants—should enable us to directly compare whether the type of task—lexical identification task versus phoneme identification task—may be at least partly responsible for the divergent results in previous studies.

2. METHOD

2.1. Participants

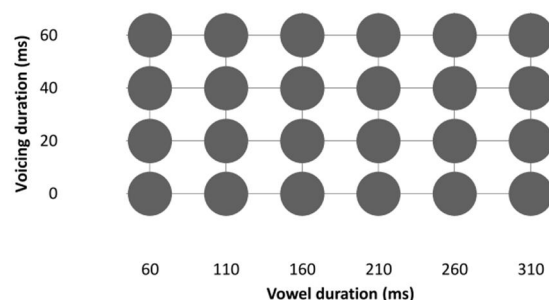
Twenty-four native speakers of Japanese aged 18 to 37 years old (mean= 21.5) and 10 native speakers of English aged 23 to 57 years old (mean = 39.6) participated in our experiment. The native Japanese listeners started learning English at school between 3 and 13 years old (mean=9.8) for an average of 11.5 years of English education. None of them had stayed in an English-speaking country for more than 1 month. The native English speakers were all from North America as in [10], but recruited in Tokyo. Their proficiency in Japanese varied greatly, from low to high proficiency.

2.2. Stimuli

The stimuli were the same as those used in [10]. They were manipulated from a /bɪd/ sample produced by a female native speaker of Canadian English. The manipulations were conducted over the voicing duration (duration of periodicity during the final stop closure) and the preceding vowel duration. The duration of voicing was gradually removed by

20ms from the endpoint, which was first set to 60ms from the vowel offset, resulting in four different variants of voicing duration. Then, vowel duration was manipulated in equal steps of 50ms, and varied from 60ms to 310ms. In total, 24 stimuli varying in vowel duration and voicing duration were created. The other acoustic cues were not modified, with F1=415Hz, F2=2163Hz, F3=3027Hz, F4=4130Hz, F5=4846Hz, the closure duration of the word final stop fixed to 100ms, and the following release burst to 35ms.

Figure 1: Stimuli used for the experiment, manipulated in vowel duration and voicing duration (adapted from [10]: 773).



2.3. Procedure

The experiment consisted of a computerized forced-choice identification task like the one used in [10]. The participants hear the stimuli and had to choose which word/sound they think they heard. Whereas in [10] they were given the choice between the entire words 'bit' and 'bid', in our experiment they were given the choice between only the final sounds 't' and 'd'. In both studies, the next stimulus was presented 1500ms after the response of the participant.

The task consisted of 1 practice block and 3 experimental blocks, with each block containing all 24 stimuli. That is, each participant had to classify a total of 96 tokens. The order of the trials was randomized within each block. While participants were not told that there was a practice block, the result of the practice block was excluded and 72 trials remained for analysis.

3. RESULTS

English listeners in the lexical identification task [10] showed significant use of both periodicity ($r = .530, p < .001$) and vowel duration ($r = .446, p < .001$), as shown in Table 1. English listeners in the phoneme identification task (current study) also showed significant use of both periodicity ($r = .591, p < .001$) and vowel duration ($r = .412, p < .001$), as shown in Table 2. This indicates that the type of task

did not affect the cue-weighting by native English listeners.

Table 1. Regression results for English listeners under the lexical identification task, cited from [10]

	B	SE B	
Constant	-.307	.039	
Periodicity	.194	.011	.530*
Vowel Duration	.107	.007	.446*

Note: Model $R^2=.479$, $*p<.001$, B = regression coefficient, SE B = standard error of B, = standardized regression coefficient

Table 2. Regression results for native English listeners under the phoneme identification task

	B	SE B	
Constant	-.989	.178	
Periodicity	.655	.050	.591*
Vowel Duration	.299	.033	.412*

Note: Model $R^2=.589$, $*p<.001$

On the other hand, Japanese listeners showed a difference depending on the type of task. The results in [10] indicate that Japanese listeners used both periodicity ($=.397$, $p<.001$) and vowel duration ($=.537$, $p<.001$) when faced with a lexical identification task, as shown in Table 3. Our results, however, indicate that Japanese listeners used only periodicity ($=.646$, $p<.001$), not vowel duration ($=.002$, $p=.96$), in the phoneme identification task, as shown in Table 4.

Table 3. Regression results for Japanese listeners under the lexical identification task, cited from [10]

	B	SE B	
Constant	-.261	.040	
Periodicity	.144	.011	.397*
Vowel Duration	.128	.007	.537*

Note: Model $R^2=.446$, $*p<.001$

Table 4. Regression results for Japanese listeners under the phoneme identification task

	B	SE B	
Constant	-.272	.128	
Periodicity	.726	.036	.646*
Vowel Duration	.001	.023	.002

Note: Model $R^2=.417$, $*p<.001$

Figure 3 shows the response time of the Japanese listeners in the lexical identification task as reported

in [9]. It indicates that Japanese listeners tend to respond faster for the extremely short or long vowels. Regression analyses conducted on the RT data split into short (60-160ms) and long (210-310ms) vowels, confirmed a small significant effect of vowel duration when the vowels are short ($=.125$, $p<.05$).

Similar separate regression analyses on the RT data of Japanese listeners in the phoneme identification task illustrated in figure 4 revealed also a significant effect of vowel duration for stimuli with short vowel duration ($=.091$, $p<.01$). This means that even though vowel duration may not have been taken into consideration in their categorical decisions, Japanese listeners had more hesitation as the vowel duration went from 60ms to 160ms.

Figure 3: Average (log-transformed) response times for the Japanese listeners in lexical identification task ([9]: 214)

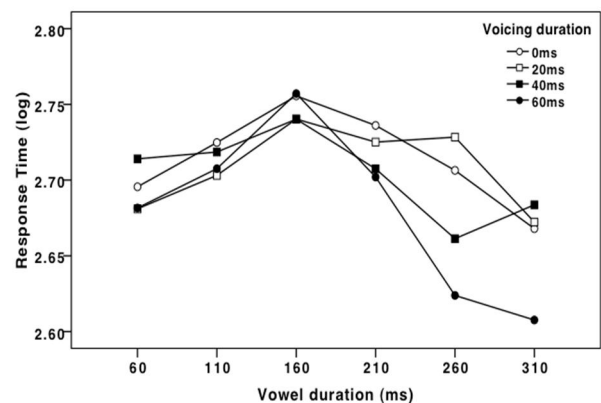
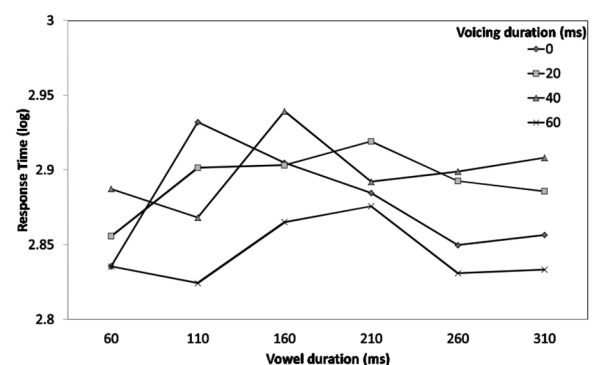


Figure 4: Average (log-transformed) response times for the Japanese listeners in phoneme identification task



In sum, despite their obvious sensitivity to vowel duration, Japanese listeners showed a different usage of vowel duration as a cue depending on the type of task, lexical identification or phoneme identification, while English listeners consistently used vowel duration in both types of task.

4. DISCUSSION

Native English listeners used vowel duration to the same extent when contrasting voiced and voiceless coda consonants irrespective of the task used: phonemic identification (current study) or lexical identification [10]. Native Japanese listeners, on the other hand, did not use vowel duration in the phoneme identification task (current study), although they used vowel duration when faced with a lexical identification task [10]. Hence, although the difference in the two types of task employed in these studies did not result in a difference in the use of vowel duration by native listeners, it did result in a difference in the use of vowel duration by L2 listeners. The question is "why?"

A possible explanation would be that, while English listeners used the vowel duration contrast as a cue to the coda voicing contrast /t/ and /d/ in both tasks, Japanese listeners applied their native vowel contrast /i/ and /i:/ to the words $\text{-bit}\emptyset$ and $\text{-bid},\emptyset$ not to the coda voicing contrast. That is, they thought the word 'bid' had a long vowel and the word 'bit' a short vowel, without associating the vowel contrast with the word final stop. If this is the case, that means that native Japanese listeners never dissociated their use of vowel duration from the vowel contrast in their L1 to perceive the coda voicing contrast in L2 as previously concluded in [10].

It is already known that a difference between types of task e.g., identification task vs. ABX task can trigger different results [4][5]. Our results indicate that the same task can also trigger different outcome depending on the choices provided to the test taker.

Another issue relates to orthographical representations. Previous studies have argued that the orthographical difference is necessary for L2 learners to lexically represent the contrast [7][11]. For example, [7] shows that Dutch learners of English could discriminate the words containing the novel contrast /ɛ/ vs. /æ/ only when they had access to the spelled form, which gives cue to the phonetic difference.

Our results, on the other hand, put into question the role of orthography in L2 acquisition. Orthography plays a negligent role in English loanwords in Japanese [8], and our results suggest it does not always play a significant role in L2 perception either. Japanese listeners appear to represent the vowels in $\text{-bit}\emptyset$ and $\text{-bid}\emptyset$ as /i/ and /i:/ respectively, in spite of the same orthographical representation, and despite the fact that they are known to also associate a vowel length contrast to

the tense-lax vowels in English, as in 'beat' and 'bit' [9][12].

In conclusion, the acoustic cues used in L1 may not be readily applied to novel phonemic contrasts in L2. The learners, however, can encode the familiar acoustic difference lexically, and that, possibly while disregarding orthographical information.

7. REFERENCES

- [1] Booij, G. 1995. *The phonology of Dutch* Oxford, England: Oxford U Press.
- [2] Broersma, M. 2005. Perception of familiar contrasts in unfamiliar positions. *J. Acoustic. Soc. Am.*, 117(6), 3890-3901.
- [3] Crowther, C. S., & Mann, V. 1992. Native language factors affecting use of vocalic cues to final consonant voicing in English. *J. Acoustic. Soc. Am.*, 92(2), 711-722.
- [4] Curtin, S., Goad, H., & Pater, J. V. 1998. Phonological transfer and levels of representation: the perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research*, 14(4), 389-405.
- [5] Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. H. 2012. Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English- L2 French acquisition. *Second Language Research*, 28(1), 5-40.
- [6] Escudero, P. R. 2005. *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Utrecht: Landelijke Onderzoekschool Taalwetenschap.
- [7] Escudero, P., Hayes-Harb, R., & Mitterer, H. 2008. Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, 36(2), 345-360.
- [8] Goulet, MJ. 2001. *L'adaptation phonologique d'emprunts lexicaux anglais en japonais et la faible influence de la graphie*. M.A. thesis, U. Laval.
- [9] Grenon, I. 2010. *The bi-level input processing model of first and second language perception*. Ph.D. thesis. U. of Victoria.
- [10] Grenon, I. 2011. Dissociable levels of speech processing in second language perception. *Proc. 17th ICPhS Hong Kong*, 771-774.
- [11] Hayes-Harb, R., & Masuda, K. 2008. Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, 24(1), 5-33.
- [12] Morrison, G. S. 2002. *Effects of L1 Duration Experience on Japanese and Spanish Listeners' Perception of English High Front Vowels*. MA thesis. Simon Fraser University.