DISSOCIABLE LEVELS OF SPEECH PROCESSING IN SECOND LANGUAGE PERCEPTION

Izabelle Grenon

Graduate School of Literature, Seijo University, Tokyo, Japan
izabelle.grenon@yahoo.ca

ABSTRACT

It is generally accepted that there are at least two levels of speech processing: a level that processes categorical (and possibly indexical) acoustic information, and a level that processes abstract information for phonemic/lexical contrasts. However, little is known about how these levels may interact to impede or facilitate the perception of non-native speech contrasts. Our study evaluates whether native Japanese speakers can take advantage of their sensitivity to vowel duration—used for an underlying vowel contrast in L1—and periodicity—used for an onset voicing contrast in L1—to perceive the English 'bit'-'bid' coda voicing contrast. Our results suggest that Japanese speakers are capable of extending their use of both vowel duration and periodicity to perceive the English coda voicing contrast in a way that is closely comparable to native English speakers. These findings are interpreted as indicating that adult L2 learners can dissociate acoustic processing from phonological processing, and capitalize on their sensitivity to acoustic cues used for phonological contrasts in L1 to perceive non-native contrasts.

Keywords: second language; perception; speech sound categories; English; Japanese

1. INTRODUCTION

Adult second language (L2) learners may encounter difficulties perceiving L2 contrasts not employed in their L1 [4, 10]. The source of L2 learners' difficulties with non-native contrasts may partly stem from their inability to efficiently process the acoustic cues critical to the L2 contrasts [9, 11]. For instance, Japanese speakers' difficulty to perceive the English /r/-/l/ contrast is attributed to Japanese speakers' lack of sensitivity to changes in the third formant (F3) around the categorical boundary critical to the /r/-/l/ contrast in English [9], presumably because changes in F3 are not relevant for any contrast in Japanese [7] (there is no /r-l/ phonemic contrast in Japanese).

On the other hand, the fact that an acoustic cue occurs contrastively in the listeners' native language (L1) is still no guarantee that contrasts using that cue will readily be perceived. For instance, despite the fact that the F2 onset is contrastively higher in the sequence [da] than in the sequence [sta] in English, when the 's' is stripped off the latter sequence, American English speakers identify both stops (i.e. in the sequences [da] and [ta], both with short VOT) as equally good instances of the phoneme /d/ when the tokens are presented through a forced-choice identification task [12]. These results indicate that listeners cannot necessarily extend the use of an acoustic cue that occurs contrastively in their L1 to categorically perceive novel speech sounds based on that cue. In the case described above, changes in F2 are presumably ignored because the presence or absence of [s] is sufficient to distinguish [da] from [sta] since the sequence *[sda] is not permissible in English. As a result, it is possible that in order for L2 learners to be able to perceive an L2 contrast—without training or instruction—the critical acoustic cue for this contrast must not only be contrastive in their L1, but must also be used to distinguish a phonological contrast in their L1. However, does that mean that an acoustic cue associated with a phonological contrast in L1 may be used to perceive *any* new contrast? In other words, can acoustic processing be dissociated from the phonological feature it is usually linked to in L1 and used to perceive a novel underlying contrast? Although various levels of speech processing have generally been recognized [3, 14, 15], it remains to be confirmed whether adult L2 learners can dissociate these levels and capitalize on their sensitivity to a contrastive acoustic distribution to perceive novel speech contrasts.

The present study investigated this question by evaluating whether Japanese speakers may use vowel duration—linked with an underlying vowel contrast in their L1—to perceive a coda voicing contrast in English, when they are given a choice between using vowel duration and using the

presence of periodicity during stop closure (or both). Since the presence of periodicity is used in Japanese and already linked with an underlying voicing contrast—albeit only in onset position—we might expect Japanese speakers to favor the use of periodicity over the use of vowel duration in order to categorize the English voicing contrast in coda position.

Native English speakers may use multiple acoustic cues to distinguish a word like 'bid' from 'bit'. The most productive have been identified as the use of vowel duration and the presence of periodicity during the coda stop closure [8]. Japanese speakers also use the presence of periodicity for a voicing contrast, where /d/ has a negative VOT and /t/ a short VOT in onset position [13]. However, no voicing contrasts are permissible in coda position in Japanese [1].

On the other hand, the use of vowel duration is very productive in Japanese and used to contrast phonemically short and long vowels [1]. It has been shown that Japanese speakers are able to apply their sensitivity to vowel duration to perceive a novel L2 vowel contrast [11], but it is still unknown if they can capitalize on the same cue to perceive a different type of underlying contrast, such as the English coda voicing contrast.

Given Japanese speakers' sensitivity to vowel duration to contrast short and long vowels, and to the presence of periodicity during stop closure to contrast voiced from voiceless stops in onset position, two hypotheses concerning the perception of the English coda voicing contrast by Japanese speakers are possible, and may have theoretical implications for L2 perception and acquisition: (1) if the linkage between acoustic processing and phonological encoding is unalterable, Japanese speakers are predicted to rely exclusively on the presence of periodicity to distinguish 'bid' from 'bit'; or (2) if the linkage between acoustic processing and phonological encoding is instead relatively loose, Japanese speakers are expected to use both periodicity and vowel duration to classify the English coda voicing contrast. The latter prediction is in line with general L1 models of speech perception that posit dissociable levels (or planes) of speech processing, such as the PRIMIR model [14], and in line with previous L2 experiments [3, 5], although this idea is still in need of further empirical support, especially from L2 data with acoustically-controlled stimuli. The current study was designed to fill this gap by evaluating whether Japanese speakers may categorize the English stop voicing contrast in coda position by relying exclusively on the presence of periodicity (used for the same voicing contrast in their L1 but in onset position), or whether they may also rely on vowel duration (used for a different contrast in their L1). English speakers were tested for comparison.

2. METHODOLOGY

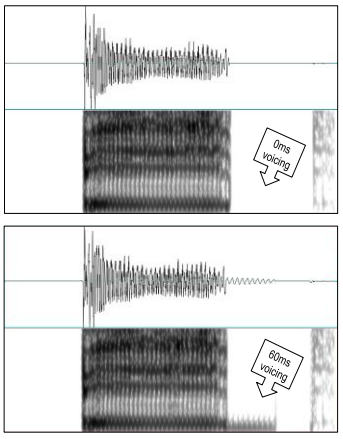
2.1. Participants

Twenty-four native speakers of Japanese recruited in Tokyo, and 24 native speakers of North American English recruited in Western Canada participated in this experiment. None of the participants reported having any known hearing impairments. The Japanese participants were from monolingual homes and had never lived abroad with the exception of one female speaker who spent 10 months in Australia.

2.2. Stimuli

Twenty-four 'bit' and 'bid' tokens were created by cross-splicing and editing portions of a natural speech sample—recorded from a female Canadian English speaker—using Praat [2].

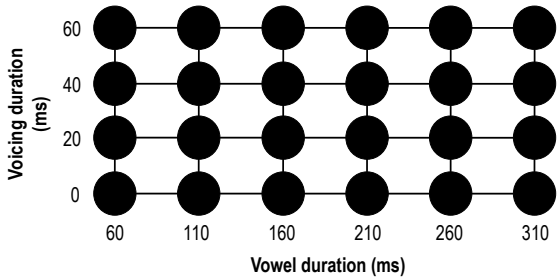
Figure 1: Test tokens with 0ms of periodicity (top) and 60ms of periodicity (bottom).



Periodicity during closure duration of the natural /bɪd/ sample recorded was first set to a duration of 60 ms starting at the vowel offset. Portions of the periodicity were gradually removed from the end of the 60 ms in segments of 20 ms, and varied from 0 ms to 60 ms. Samples of manipulated tokens differing in terms of voicing duration (periodicity) are provided in Figure 1. Vowel duration was then varied in equal steps of 50 ms, from 60 ms to 310 ms, to create four continua varying in the duration of periodicity

during stop closure and vowel duration, as schematized in Figure 2.

Figure 2: Tokens used for the experiment, which vary in terms of vowel duration and duration of periodicity during word-final stop closure.



The remaining acoustic cues were held constant across all 24 test tokens: F1 = 415 Hz, F2 = 2163 Hz, F3 = 3027 Hz, F4 = 4130 Hz and F5 = 4846 Hz. The closure duration of the word-final stop consonant was fixed to 100 ms, and its following release burst to 35 ms.

2.3. Procedure

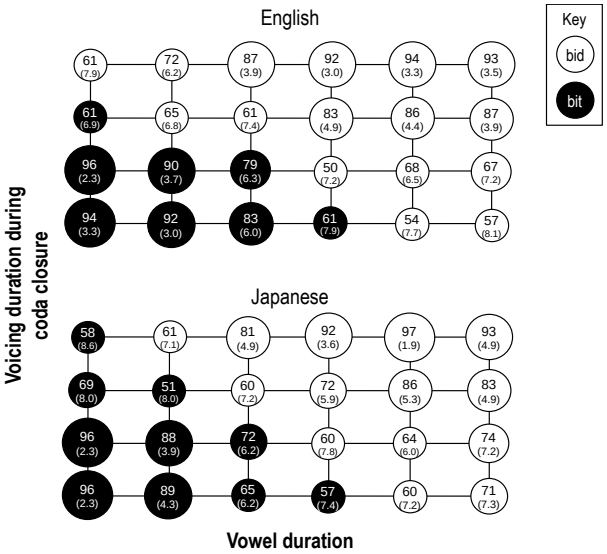
A computerized forced-choice identification task was used for this experiment: participants listened to one word presented in isolation and had to select which word, out of two choices ('bit' or 'bid'), the word they heard corresponded to. A long interval of 1500ms followed the participant's response before presentation of the next stimulus.

Each participant completed a practice block of 24 trials with each of the possible tokens presented once in a random order. After completing the practice block, the experimental session consisted of three blocks of the 24 tokens (for a total of 72 test trials) with the order of trials randomized within each block. The experiment lasted about 10 minutes, and was part of a larger experiment.

3. RESULTS

The averaged identification patterns for the English and Japanese groups are very similar, as shown in Figure 3 below. In this figure, a white circle corresponds to a token identified mainly as 'bid' and a black circle to a token identified in most cases as 'bit'. English and Japanese speakers appear to have taken into consideration vowel duration and the presence of periodicity during the coda stop closure to categorize the manipulated stimuli: tokens with a short vowel and short voicing duration are generally identified as 'bit' (left bottom rows), whereas tokens with a longer vowel and longer voicing duration (right columns and top rows) are generally identified as 'bid' by speakers of both language groups.

Figure 3: Averaged identification of tokens as either 'bit' or 'bid' across English (top) and Japanese (bottom) listeners. The size of each circle corresponds to its identification frequency: large circles indicate higher identification percentages. The shading of the circle indicates the most frequently identified category: white for 'bid' and black for 'bit'. The number within each circle indicates the identification percentage for the most frequently identified category, with standard error in parentheses.



A mixed-design ANOVA reports a small but significant effect of native language (L1) on the use of periodicity, $F(2.09, 140) = 4.67$, $p < .01$. On the other hand, the interaction between L1 and vowel duration was not significant ($F[3.15, 138] = 1.75$, $p > .05$); nor was the three-way interaction between L1, vowel duration, and periodicity ($F[12.45, 128] = .938$, $p > .05$), indicating that Japanese speakers used vowel duration in a way comparable to native English speakers. Although Japanese speakers did not use periodicity in a way statistically comparable to native English speakers, the identification results suggest that this cue still played an important role in their categorical decision. Hence, multiple regressions were performed on the English and Japanese data (separately) to evaluate the relative use of each independent variable—periodicity and vowel duration—for categorization of the coda contrast.

The effect of periodicity and vowel duration predicts 48% of the identification responses for English speakers ($R^2 = .479$). As summarized in Table 1, the effect of periodicity was slightly greater in the English data ($\beta = .530$, $p < .001$) than the effect of vowel duration ($\beta = .446$, $p < .001$).

Table 1: Regression results for English speakers.

	B	SE B	β
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, β = standardized regression coefficient.

The combined effect of periodicity and vowel duration explains about 45% of the Japanese identification results (R^2 = .446). Unlike the participants in the English group, Japanese participants appear to use vowel duration (β = .537, p < .001) to a greater extent than periodicity (β = .397, p < .001).

Table 2: Regression results for Japanese speakers.

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

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

In sum, native Japanese speakers were able to categorize the coda stop voicing contrast, using information provided by both vowel duration and periodicity during coda stop closure in a way closely comparable to native English speakers.

4. DISCUSSION

The goal of the present study was to assess whether Japanese learners of English could extend their use of an acoustic cue already used for a phonological contrast in their L1 (vowel duration) to perceive a novel L2 contrast (English coda voicing contrast), when more than one possible cues are available. Results of this experiment suggest that L2 learners can extend their sensitivity to acoustic cues used to perceive phonemic contrasts in their L1 to perceive novel L2 contrasts based on that cue, and this without any training or instruction about which cue is relevant for categorization of the L2 sounds (see also [5]). These results are interpreted as providing evidence for (at least) two dissociable levels of speech processing in L2 perception (hypothesis 2 in Introduction), since the Japanese speakers in our study were able to dissociate the use of vowel duration from the phonological vowel contrast as used in their L1 and apply their sensitivity to vowel duration to distinguish the consonant coda voicing contrast in English. And this, despite the fact that periodicity, a cue already used for a voicing contrast in their L1, was also available to them. The findings presented here concur with recent multi-level models of L2 speech perception (e.g. [6]) by providing evidence that L2 learners can extend the use of an acoustic cue contrastive in their L1 to readily perceive (i.e. without training or instruction) a non-native speech contrast, if—and based on previous findings possibly *only if*—this cue is already used to perceive a phonological contrast in their L1.

5. REFERENCES

[1] Akamatsu, T. 1997. *Japanese Phonetics: Theory and Practice.* Newcastle: Lincom Europa.

[2] Boersma, P., Weenink, D. 2007 *Praat: Doing phonetics by computer* [computer program].

[3] 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. *Sec. Lang. Research* 14, 389-405.

[4] Flege, J.E. 1993. Production and perception of a novel, second-language phonetic contrast. *J. Acoust. Soc. Am.* 93, 1589-1608.

[5] Flege, J.E., Hillenbrand, J. 1986. Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. *JASA* 79(2), 508-517.

[6] Grenon, I. 2010. *The Bi-Level Input Processing Model of First and Second Language Perception*. Ph.D. dissertation. U of Victoria. *http://hdl.handle.net/1828/2907*.

[7] Guenther, F.H., Bohland, J.W. 2002. Learning sound categories: A neural model and supporting experiments. *Acoustical Science and Technology* 23(4), 213-221.

[8] Hogan, J., Rozsypal, A. 1980. Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *J. Acoust. Soc. Am.* 67, 1764-1771.

[9] Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., Siebert, C. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47-B57.

[10] Logan, J.S., Lively, S.E., Pisoni, D.B. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *J. Acoust. Soc. Am.* 89, 874-886.

[11] 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.

[12] Pegg, J.E., Werker, J.F. 1997. Adult and infant perception of two English phones. *J. Acoust. Soc. Am.* 102, 3742-3753.

[13] Shimizu, K. 1996. *A Cross-Language Study of Voicing Contrasts of Stop Consonants in Asian Languages*. Tokyo: Seibido Publishing.

[14] Werker, J.F., Curtin, S. 2005. PRIMIR: A developmental framework of infant speech processing. *Lang. Learning and Dev.* 1(2), 197-234.

[15] Werker, J.F., Logan, J.S. 1985. Cross-language evidence for three factors in speech perception. *Perception & Psychophysics* 37 (1), 35-44.