L1 ALLOPHONES AND L2 SOUND PERCEPTION

Izabelle Grenon

The University of Tokyo
grenon@boz.c.u-tokyo.ac.jp

ABSTRACT

The bi-level input processing model posits two levels of speech sound processing. In this model, context-bound allophones are encoded separately at a lower level of processing. If this is the case, second language learners should exhibit some facilitation effect when perceiving non-native phonemes that are used as context-bound allophones in their first language. Using a cue-weighting design, the current study tested the hypothesis that Canadian French listeners should be able to apply their sensitivity to spectral changes in the high front tense-lax vowel allophones in their first language to perceive the high front tense-lax vowel phonemes in English. Our results demonstrate that most of the Canadian French listeners could perceive the English vowel contrast in a way similar to North American native English listeners. We further discuss possible explanations as to why some could not, and why a previous study with Spanish listeners did not demonstrate such facilitation effect.

Keywords: second language, speech perception, levels of processing, allophonic contrasts

1. INTRODUCTION

Recent models of speech perception posit different levels of processing to capture the fact that variations in tasks and task conditions may tap differentially into levels of sound representations [3, 4, 7]. Allophones, in particular, are hypothesized in the bi-level input processing (BLIP) model to be encoded separately at an early stage of speech processing and inter-connected only at a higher level [4]. If this is the case, listeners should be able to perceive allophonic contrasts in their first language (L1) under conditions leading to acoustic perception, and second language (L2) listeners should possibly exhibit some facilitation effect when perceiving L2 contrasts corresponding to the allophonic variants in their L1.

Native English speakers tend to rate [da] and the voiceless unaspirated [ta]—an allophone of the voiceless aspirated phoneme /t/ extracted from the sequence /sta/—as equally good instances of /da/ in a rating task, although they are able to discriminate the same sounds above chance level in an AX discrimination task [14]. This suggests that native English listeners cannot distinguish the two unaspirated sounds at a phonemic level, but can discriminate them at a surface level of processing.

A training experiment was conducted evaluating whether English listeners could improve their perception of a contrast when presented with a contrastive distribution of those sounds [9]. The experiment used comparable sounds as in the previous study—i.e. voiced /d/ (as in *day*) and voiceless unaspirated /t/ (as in *stay*)— but here the English learners were told these sounds belong to an "unknown" language. With only 9 minutes of passive exposure to the sounds presented in a contrastive distribution—i.e. without having to perform any identification or discrimination task throughout the training phase—native English listeners could improve their perception of these sounds as assessed through an AX discrimination task.

Similarly, French and English native listeners' training with the Thai 3-way stop contrast resulted in English but not French listeners being able to perceive the stop aspirated and unaspirated contrast in an ABX task, presumably because English but not French listeners are sensitive to these variants as allophones in their L1 [2]. However, in the same study, there was no facilitation effect for the English group when the task involved a picture identification task using minimal-pairs with the same sound contrast.

The fact that the L2 sounds are allophones in the L1 is no sure guarantee, however, that a facilitation effect will occur, even when the allophones are in complementary distribution in the listeners' L1. Spanish listeners did not show any advantage over native Japanese listeners in discriminating the English voiced alveolar stop as in 'day' and voiced interdental fricative as in 'they' in an AXB task despite the fact that Spanish speakers generally produce a voiced interdental fricative as an allophonic variant of /d/ in intervocalic context [13].

Accordingly, whether listeners may more easily perceive L2 contrasts that are allophonic in their L1 may depend on the type of task and the targeted sound contrast. To clarify this issue, it may be necessary to understand whether the L2 listeners are attuned to the same acoustic cues that native listeners rely on when perceiving the target sounds.

The current paper evaluates how L2 learners use the acoustic cues available in an L2 vowel contrast when these vowels are context-bound allophones in their L1, and when using an identification task rather than a discrimination task. An AX task is generally used to evaluate the ability of listeners to discern small acoustic differences, whereas the ABX or AXB task evaluates listeners' ability to ignore irrelevant acoustic differences for phonemic categorization. The cue-weighing task employed in this paper should instead shed light on which acoustic cues, if any, the L2 listeners rely on to classify the L2 contrasts when the crucial acoustic cues are used contrastively at an allophonic level in their L1.

2. THE BLIP MODEL AND PREDICTIONS

The Bi-Level Input Processing (BLIP) model [4] posits 2 levels of speech processing (besides lexical encoding): A neural mapping level and a phonological level. In-line with previous neural models of sound processing [5, 6, 16], the BLIP model posits that neural maps are affected by the statistical distribution of acoustic cues in input, where a contrastive distribution should trigger the formation of contrastive neural maps. These maps are in turn associated with abstract, phonemic representations.

During the first year of life, infants are sensitive to the statistical distribution of acoustic cues in input [10, 11]. Accordingly, the BLIP model suggests that if the distribution of allophones is sufficiently contrastive in input, infants should develop distinct neural maps for each statistically contrastive acoustic contrast. As language development progresses and it becomes clear that these cues are not contrastive at a higher level, this distinction becomes irrelevant. However, the BLIP model posits that the neural maps remain separated potentially throughout one's lifespan, though they become associated with the same underlying representation at the phonological level.

Canadian French speakers (CF), as opposed to European French speakers, are known to produce a high front tense vowel in open syllable (e.g. 'lit' [li̠] *bed*) and a lax allophone in closed syllable (e.g. 'lime' [lɪm] *lime*) (cf. [8] for more examples). These vowels are acoustically closely comparable to the English high front tense-lax vowel contrast, as in 'beat-bit'. Provided that the context-bound variation in Canadian French exhibits a sufficiently contrastive distribution along the first (F1) and second formant (F2) dimension, the BLIP model predicts that CF listeners should have developed separate neural maps based on spectral differences to process these vowels during infancy, although these maps came to be associated with the same underlying /i/ vowel at the phonological level. That is, the BLIP model predicts that CF listeners should be able to distinguish the English vowel contrast, at least at the neural mapping level. The following experiment evaluates this hypothesis by testing perception of the English vowel contrast by CF listeners as well as North American English listeners.

The acoustic cues manipulated for the current cue weighting experiment are formants and vowel duration: The former to verify whether CF listeners can rely on spectral differences that are only used at an allophonic level in their L1, and the latter to serve as a distracter. L2 listeners who cannot distinguish the English vowels based on spectral contrast often rely on duration instead ([4, 12] for Japanese; [17] for Mandarin and Cantonese). If vowel duration is not contrastive in their L1, it is also possible that L2 listeners will use neither cue ([12] for Spanish).

3. METHOD

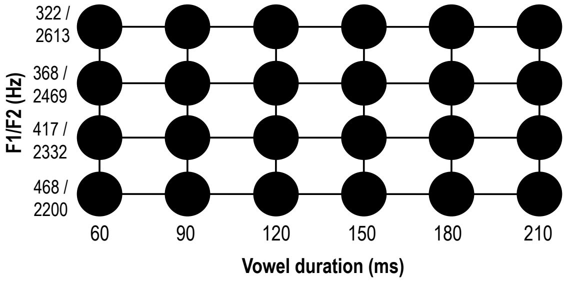
3.1. Participants

Twenty-four Canadian French (CF) listeners recruited in Québec, Canada, and 24 North American English (NAE) listeners recruited in Western Canada participated in this experiment. None of the participants reported having any known hearing impairment. The CF participants were from monolingual homes and had never lived abroad with the exception of one participant who had spent five weeks in an English immersion program. The CF listeners were between 17 and 29 years old (mean 21.3); started studying English at school between 8 and 12 years old (mean 9.6); and had completed on average 8.9 years of education in the English language. The NAE listeners were between 18 and 30 years old (mean 20.4).

3.2. Stimuli

Twenty-four 'bit' and 'beat' tokens were created by cross-splicing and editing portions of a natural speech sample—recorded from a female Canadian English speaker—using Praat [1]. First, a 'bit' sample was modified to set the F1 and F2 to 468Hz and 2200Hz respectively. The F1 was subsequently lowered and F2 increased in steps of 50 Mel [15], yielding four spectrally different vowels: F1(468Hz)/F2(2200Hz), F1(417Hz)/F2(2332Hz), F1(368Hz)/F2(2469Hz), F1(322Hz)/F2(2613Hz). Vowel duration was then varied in equal steps of 30 ms, from 60 ms to 210 ms, to create four F1/F2 continua each varying in vowel duration, as schematized in Figure 1. For all 24 tokens used for the experiment, F3 was set to 3099Hz, F4 to 4115Hz, and F5 to 5000Hz. The formant transitions in word-initial and word-final positions were not manipulated, nor were any of the formant bandwidths or pitch contours. The closure duration in the production of the final consonant was fixed to 100ms and the burst release to 130ms for all tokens.

Figure 1: Duration and F1/F2 values for the vowels in the bit/beat tokens used for the experiment.



3.3. Procedure

For this experiment, participants completed a computerized two-choice identification task: they listened to one word presented in citation context and had to select which word they thought they heard ('bit' or 'beat'). An interval of 1500ms followed each participant's response before presentation of the next test token.

Each participant completed a practice block of 24 trials with each of the possible tokens presented once in a random order. After completing this practice block, the experimental session consisted of three blocks including each of the 24 tokens (for a total of 72 test tokens) with the order of tokens randomized within each block. The experiment lasted about 5-10 minutes, and was part of a larger experiment.

3. RESULTS

The averaged identification results for the NAE and CF listeners are very similar, as shown in Figure 2. In this figure, a white circle corresponds to a stimulus identified in most cases as 'beat' and a black circle to one identified as 'bit'. Tokens containing vowels with high F1 and low F2 are generally identified as 'bit' by listeners of both groups while tokens containing vowels with low F1 and high F2 are identified as 'beat'. That is, both CF and NAE listeners appear to rely mainly on spectral changes, rather than changes in vowel duration.

To evaluate the exact use of formants and vowel duration, we conducted regression analyses on the two groups separately. Table 1 presents the results for the NAE listeners. Changes in formants and duration account for 72% of the results in this model (R^2= .723), with NAE listeners relying on changes in formants more (β = .814, p < .001) than on changes in vowel duration (β = .247, p < .001), though both cues are used to a statistically significant level.

Figure 2: Averaged identification of tokens as either 'beat' or 'bit' across English (top) and French (bottom) listeners. The size of circle represents its identification frequency in percentage, with each value within each circle with standard error in parentheses. The shading (black or white) indicates the most frequently identified category.

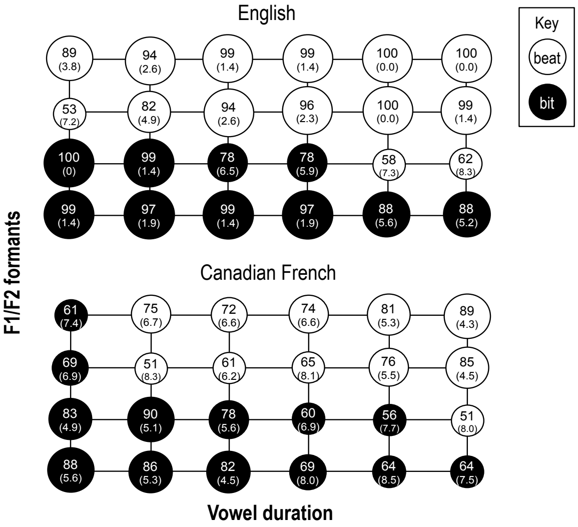


Table 1: Regression results for English listeners.

	B	SE B	β
Constant	-.521	.032	
Formants	.333	.009	.814*
Duration	.066	.006	.247*

Note: Model R^2 = .723, *p < .001, B = regression coefficient, SE B = standard error of B, β = standardized regression coefficient.

Similarly, CF listeners relied on both cues as shown in Table 2. However, the use of the two cues account for only 34% of the results here (R^2 = .344). Although they use formants to a larger extent than vowel duration, like NAE listeners, the CF listeners appear to rely on duration (β = .338, p < .001) more than NAE listeners, and on formants less (β = .479, p < .001).

Table 2: Regression results for French listeners.

	B	SE B	β
Constant	-.237	.043	
Formants	.172	.012	.479*
Duration	.079	.008	.338*

Note: Model R^2 = .344, *p < .001

However, some CF listeners may rely on formants only, while others rely on duration only. We looked at individual data by using a mathematical criterion to separate them according to the acoustic cue they appear to most rely on. This mathematical criterion is the bias-ratio as introduced and justified in [4]. Based on this criterion, it was found that NAE listeners exhibit two possible patterns: 19 used mainly spectral changes (formants), while 5 used both spectral and vowel duration changes (formants+duration), as reported in Table 3. Twelve of the CF listeners also used mainly formants, 3 used formants and duration, while 5 used mainly vowel duration and 4 exhibited no obvious bias. Hence, the majority of the CF listeners (i.e. 12+3 = 15/24) had a pattern of identification comparable to that of NAE listeners.

Table 3: Number of NAE and CF listeners exhibiting each of the possible bias patterns.

	English (N=24)	French (N=24)
Formants	19	12
Formants+duration	5	3
Duration	0	5
No bias	0	4

4. DISCUSSION

This study evaluated whether CF listeners could capitalize on their sensitivity to spectral differences in the [i-ɪ] allophonic contrast in their L1 to perceive the English /i-ɪ/ phonemic contrast. The results of the current cue-weighting experiment suggest that most of the CF participants could rely on spectral differences to perceive the tense-lax English vowel contrast, although not all of them did. The fact that not all of our CF participants used changes in formants may simply be due to a lack of awareness as to which cue they should pay attention to, or more critically, to individual differences in perceptual sensitivity to the spectral cues.

To clarify this issue, we plan to conduct a training experiment with CF listeners, in which we tell the participants which cue they should attend to. If awareness is sufficient, all of the tested CF should be able to use formants post-training with the English tense-lax vowels. For comparison, we would like to test and train European French listeners. Since the latter lack the allophonic contrast in their French variety, the BLIP model predicts that they should not show any facilitation effect. Hence, before training, most of them should be unable to use formants, and training should yield inferior results than for the CF listeners.

Why did the CF listeners in this study show a facilitation effect to perceive an L2 contrast presumably because this contrast is used at an allophonic level in their L1, while the Spanish speakers in [13] did not show such facilitation effect? Besides the fact that the tasks used were different and the cues on which the Spanish speakers relied for their choices in the AXB task were unclear, the allophones are also quite different. The spirantization of /d/ in Spanish results from a co-articulation effect and therefore may not be encoded by separate neural maps as posited in the BLIP model. In any case, revisions of this model may be necessary to account for the fact that not all context-bound allophones may facilitate the perception of L2 phonemes.

5. REFERENCES

[1] Boersma, P., Weenink, D. 2007. Praat: Doing phonetics by computer. <http://www.praat.org/>

[2] 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. Res., 14*(4), 389-405.

[3] Escudero, P. (2005). Linguistic perception and second language acquisition. Ph.D. dissertation. Utrecht:LOT.

[4] Grenon. I. 2010. The bi-level input processing model of first and second language perception. Ph.D. dissertation. U of Victoria.

[5] Guenther, F.H., Bohland, J.W. 2002. Learning sound categories: A neural model and supporting experiments. *Acoust. Sc. Tech., 23*(4), 213-221.

[6] Guenther, F.H., Gjaja, M.N. 1996. The perceptual magnet effect as an emergent property of neural map formation. *J. Acoust. Soc. Am., 100*, 1111–1121.

[7] Key, M.P. 2012. Phonological and phonetic biases in speech perception. Ph.D. dissertation. U of Mass.

[8] Martin, P. 1996. *Éléments de phonétique avec application au français*. Sainte-Foy: Presses U. Laval.

[9] Maye, J., & Gerken, L. 2000. Learning phonemes without minimal pairs. *Proc. BUCLD*, *24*(2), 522-533.

[10] Maye, J., Weiss, D. 2003. Statistical cues facilitate infants' discrimination of difficult phonetic contrasts. *Proc. BUCLD, 27* (2), 508-518.

[11] Maye, J., Werker, J. F., Gerken, L. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*,82(3), B101-B111.

[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 U.

[13] Muñoz Sánchez, A. 2003. The effect of phonological status on the acquisition of new contrasts: evidence from Spanish and Japanese L2 learners of English. Ph.D. dissertation. U of Cal, San Diego.

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

[15] Stevens, S.S., Volkmann, J., Newman, E.B. 1937. A scale for the measurement of the psychological magnitude pitch. *J. Acoust. Soc. Am., 8*, 185-190.

[16] Sussman, H.M. 1986. A neuronal model of vowel normalization and representation. *Brain and Lang., 28*, 12-23.

[17] Wang, X. Munro, M. J. 2004. Computer-based training for learning English vowel contrasts. *System, 32*, 539-552.