The Perception of Stops by Thai Children and Adults

Patcharee Imsri and William J. Idsardi
Walailak University and University of Delaware

1. Introduction

One of the cross-linguistic phonological differences that has been widely investigated in numerous studies is the voicing contrast (e.g., Lotz, Abramson, Gerstman, Ingemann & Nemser 1960; Caramazza, Yeni-Komshian, Zurif & Carbone 1973; Williams 1977; Flege & Eefting 1987a; Bohn & Flege 1993; Kuijpers 1996; Werker & Pegg 1997). Many of these studies have focused on second language (L2) perception and production of non-native stop categories that do *not* occur in learners' native phonological systems. Other studies focus on L2 perception and production of stop categories that are realized differently phonetically in a learner's native language (Strange et al. 1998).

The current study further investigates how L2 learners perceive non-native voicing contrasts. However, while previous research has focused on learners whose L1 has *fewer* stop categories than those that exist in a target language, the current study assesses the performance of learners whose L1 has *more* stop categories than the L2. Because of this wider inventory in the learners' native stop system, these learners are not expected to have any difficulty learning the target stop categories. It would be sufficient for them to just use their L1 system to categorize L2. The purpose of this study is, however, to investigate whether learners *do* reorganize their phonological inventory when they learn a second language, although there is no need to do so. The second language learners who are investigated in this study are Thai learners of English.

According to Lisker & Abramson (1964), stops in Thai are phonologically realized with three categories of contrast: long-lead, short-lag and long-lag voicing. These three categories are approximately the same as the three total phonetic realizations of English. However, English is a two-category language. Some native speakers of English produce the two stop categories in word-initial position as voiceless unaspirated (short-lag voicing) and voiceless aspirated (long-lag voicing). However, other English speakers produce the two categories as prevoiced (long-lead voicing) and voiceless aspirated (long-lag voicing).

---

*This study is part of the Ph.D dissertation by the first author (Imsri 2002) under the supervision of the second author. We wish to thank all the children and adults who participated in this study. We also would like to thank Cynthia Brown, Colin Phillips, Louis Arena, Irene Vogel, Benjamin Bruening and Tom Purnell for their valuable comments, suggestions and encouragement.

Since the two categories of stops in English are also phonologically contrastive in Thai, Thai learners are expected to be able to perceive and produce English stops with no difficulty. However, since Thai has three degrees of voicing, while English has two, the main question that this study aims to investigate is how these different patterns affect Thai learners' perception of English stops. They are asked to listen to the three phonetic realizations of English stops: prevoiced, voiceless unaspirated and voiceless aspirated. Results will indicate if Thai learners show two or three categories of stops in their perception of English stops.

2. Participants

Thai children and adult learners of English participated in this study. The child groups consisted of Thai primary school students in Thailand: 2nd, 5th and 6th graders, ranging in age from 8 to 12 years. The Thai adults are graduate students with ages ranging from 23 to 37 years. They have resided in the United States for more than three years. Native speakers of English, ranging in age from 19 to 40 years, also completed the tests as a control group.

3. Stimuli and Procedure

Minimal triplets of words and nonce words of the form /C(L)V…/ were created from longer forms in the following way. The target sound, C, was a labial or velar stop. Since the voiceless unaspirated [p k] do not exist in word-initial position in English, these cases were extracted acoustically from words being with an /sC-/ onset cluster, such as "sport". Although it is possible to find English words that begin with voiceless aspirated and voiced stops, to avoid any effect of editing on the stimuli, words beginning with /sVC-/ such as "support" and "subborn" were constructed and the initial syllable was removed. In additions, some stimuli also contained different final sounds, which were also removed. So, from the recorded items "sport", "support" and "subborn" the test items [por], [p^hor] and [bor] were created respectively. Full details and materials can be found in Imsri (2002).

4. Experiment 1

4.1 Discrimination Task

The discrimination experiment was conducted to investigate whether Thai and English participants could perceive the contrast between: (1) English voiced stops [b g] and voiceless unaspirated stops [p k]; and (2) English voiceless aspirated stops [p^h k^h] and voiceless unaspirated stops [p k].

4.1.2 Discrimination Results

Figure 1 illustrates d' scores obtained for performance on discrimination of the voiced/plain and the plain/aspirated contrasts by five groups of participants:

Thai 2nd, 5thand 6th graders, Thai adults and English adults. Results for the labial and velar stop categories are shown in (a) and (b), respectively.

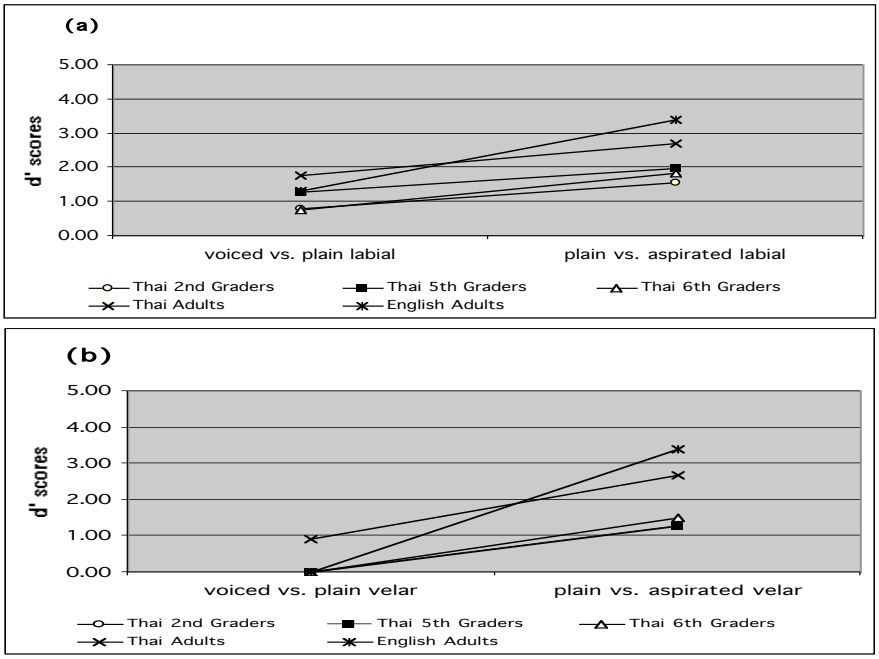


Figure 1. Mean d' scores for performance in the discrimination task: (a) labial stop categories (b) velar stop categories

A repeated measure ANOVA was conducted to determine (1) if participants from each group differed in their discrimination of the voiced/plain and the plain/aspirated contrasts; (2) if these two contrasts yielded different judgments; and (3) if distinct results were obtained for the labial and velar stop categories. The d' scores were examined in a 5 (group) x 2 (contrast-type) x 2 (place of articulation) ANOVA with repeated measures on the contrast-type and place factors. Results showed a significant main effect of group [$F(4, 59) = 9.948$, $p<.0001$], place of articulation (labial vs. velar) [$F(1, 59) = 23.427$, $p<.0001$] and contrast-type (voiced/plain vs. plain/aspirated) [$F(1, 59) = 191.428$, $p<.0001$]. No significant interactions were found between place of articulation and group [$F(4, 59) = 1.435$, $p= 0.234$], or between place of articulation and contrast-type [$F(1,59) = 2.007$, $p= 0.162$]. This suggests that participants from all five groups performed in the same way for both labial and velar stop categories. However, there was a significant interaction between contrast-type and group [$F(1,4) = 8.903$, $p<.0001$]. *Post-hoc* Bonferroni/Dunn tests revealed that the Thai childrens' d' scores were significantly lower than both the Thai and the English adults' ($p< .0001$). No significant difference was found between the two adult

groups or among the three groups of children. In addition, participants from all five groups performed significantly better in their discrimination of the plain/aspirated contrast than in the voiced/plain contrast ($p < .0001$).

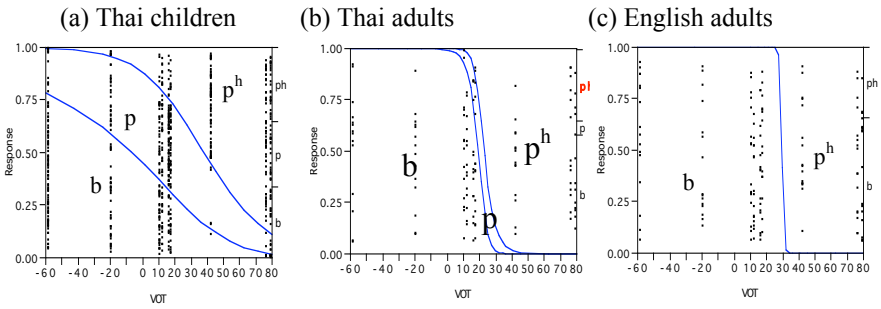
Since labial stops in Thai have three categories, but velar stops have only two categories as in English, only the judgments from the labial stop stimuli were looked at to determine whether the Thai adults and children showed two or three categories of stops in their discrimination judgments. As illustrated in Figure 1, English and Thai adults clearly showed higher d' scores in the plain/aspirated labial contrast than the voiced/plain labial contrast. This suggests that they perceived the contrast between the voiceless unaspirated and the voiceless aspirated labial stops, but did not respond to the difference between the voiceless unaspirated and the voiced labial stops. Hence, both English and Thai adults clearly showed two categories of stops in their discrimination judgments of the English labial stop stimuli. For the child groups, although their d' values for the plain/aspirated labial contrast were higher than for the voiced/plain labial contrast, these values were still very low. These children were like the adults only in their judgments of the velar stops. Their d' values for the plain/aspirated velar stops were clearly higher than for the voiced/plain velar stops. However, for their judgments of the labial stops, their d' values for the plain/aspirated contrast were not much higher than for the voiced/pain labial contrast. Hence, it is inconclusive from the discrimination results whether the Thai children groups showed two or three categories in their judgments of the labial stops. Their performance was, therefore, further investigated in the identification task.

4.2 Identification Task

All the stimuli were of the shape /C(L)V-/ and were the same stimuli used in the discrimination test. The experiment was conducted slightly differently for the Thai and English groups. For the Thai speakers, the choices for /p p^h b k k^h/ were given in Thai letters, since there are no English letters that represent the distinction between aspirated and unaspirated stops. The choices were: r [p^h], [[b], x [p], 8 [k^h], and d [k]. However, /g/ was given in English orthography since there is no appropriate letter in the Thai script. For the control group of English speakers, since there is no phonological distinction between aspirated and voiceless unaspirated stops in English, only four choices were given to the English speakers: 'p', 'b', 'k' and 'g'. The experiment was done on a Macintosh computer. Stimuli were presented through headsets. Participants were asked to identify the sound of the initial consonant of each word. They supplied their answers by pressing the appropriate button.

4.2.2 Identification Results

The logistic regression models of labial stop classification by VOT values are shown in Figure 2.

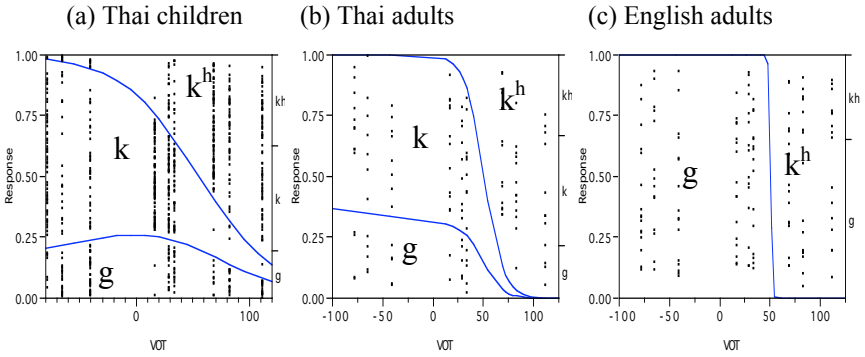


Logistic Likelihood Ratio Tests Table for Response VOT

Group	DF	Chi-square	P-Value
Thai children	2	368.45	< .0001
Thai adults	2	149.07	< .0001
English adults	1	180.85	< .0001

Figure 2. Logistic fit of English labial response by VOT

The logistic regression models for the identification of the English velar stops are shown in Figure 3.



Logistic Likelihood Ratio Tests Table for Response VOT

Group	DF	Chi-square	P-Value
Thai children	2	276.914	< .0001
Thai adults	2	114.116	< .0001
English adults	1	180.015	< .0001

Figure 3. Logistic fit of English velar responses by VOT

Figure 2 indicates that Thai children identified English stops with three categories of voicing constrast. The three large regions of 'b', 'p' and 'p^h' responses shown in the model for the Thai children suggest that their

classification of the stop categories was dependent on the VOT values of the stimuli. Stops with long-lead, short-lag and long-lag voice onset time were identified, respectively, as voiced /b/, voiceless unaspirated /p/ and voiceless aspirated /p^h/. In contrast, English adults showed two categories of stops in their responses, since they only had two choices. English speakers identified both long-lead and short-lag labial stops as voiced /b/, but categorized long-lag stops as voiceless aspirated /p^h/. As for the Thai adults, the pattern of their stop identification is more similar to that of the English adults than to that of the Thai children. Like the English speakers, Thai adults generally identified stimuli with short-lag VOT values as voiced /b/ rather than voiceless unaspirated /p/. This is illustrated by the very small region of 'p' responses for short-lag stimuli shown in the regression model for the Thai adults. In addition, the regression slopes are shallower for the Thai children, indicating that their category boundaries are less sharp than the adults'.

The logistic regression models of velar stop classification shown in Figure 3 suggest that unlike the results found in labial stop identification, Thai adults responded more similarly to the Thai children than to the English adults. Although there are only two velar stop categories in Thai, both children and adult Thai learners of English showed three categories of velar stops in their responses: voiced /g/, voiceless unaspirated /k/ and voiceless aspirated /k^h/. Both groups of Thai speakers identified English long-lag velar stimuli as /k^h/, as did the native speakers of English. However, the graphs for /k^h/ responses by Thai and English adults are much steeper than that of the Thai children, again indicating less sharp boundaries for the children. For long-lead and short-lag English velar stops, although Thai speakers sometime gave /g/ responses in their judgments, the percentage of /k/ responses was higher.

The flat slope across a large portion of the VOT continuum for the k/g boundary indicates that the choice was being made based on a factor other than VOT. Thus, it is likely that the /g/ responses reflect the choices given in the task, and should be considered an artefact of the experimental task, which mixed English and Thai orthographies in this case. Since the English stimuli had a three-way phonetic contrast, Thai speakers were given three choices for their velar judgments. While both voiceless unaspirated /k/ and voiceless aspirated /k^h/ could be presented in Thai letters, the voiced /g/ could not, as this sound does not exist in Thai. Therefore, when Thai children and adults saw different choices for [k] and [g], this might have led them to think that some sounds must be identified as /k/ and some as /g/ (or that some should be identified as "Thai" and others as "English"/"Not Thai"). Therefore, the presence of the three choices of velar stops might account for the three categories of voicing contrast found in Thai children and adults' identification of English voiced stops.

The findings of Experiment 1 have shown that L2 learners with limited exposure to English exhibit different patterns of speech perception from those learners who have more experience in English. While the Thai children showed three categories of stops, the Thai adults showed two categories of stops in their responses, as did the English speakers.

Before we can conclude that the Thai adults respond to the English stops like the native speakers because they have an L2 grammar that is different from their L1, we need to show that they perform differently when they listen to the Thai and English stop stimuli. The Thai adults' perception of English stops was then further investigated in Experiment 2, which was also a forced-choice identification task..

5. Experiment 2

In this experiment, a series of conditions were constructed to further investigate how language context shifts the perception of stop categories. There were five experimental conditions to test whether the Thai adults alter their responses according to the language they are listening to. That is, if they know that there are three categories of stops in Thai but two in English, they should show language set effects and respond with three and two categories of stops to the Thai and English stop stimuli, respectively. The Thai and English stimuli were presented in the following conditions:

Condition 1: English and Thai Stimuli are presented in *separate* trials.
Condition 2: English and Thai Stimuli interspersed within the *same* trial.
Condition 3: Same as Condition 2, but all the stimuli are shortened until the language of the stimuli is unidentifiable.
Condition 4: Same as Condition 1, but English stimuli are produced by the same Thai speaker.
Condition 5: Same as Condition 2, but both English and Thai stimuli are produced by the same Thai speaker.

5.1 Results

As shown in Table 1, below, the Thai participants perceive English and Thai stops with two and three categories of voicing contrast when they listen to English and Thai stimuli, respectively. However, the results from condition 3 (shortened stimuli) indicate that Thai participants showed neither the Thai nor the English pattern of responses. They perceived both Thai and English [p] as voiceless unaspirated roughly 50% percent of the time, and perceived them as voiced roughly 50% of the time. This is consistent with a strategy of randomly shifting back and forth between Thai and English perception modes. The English participants, since they did not have knowledge of Thai stops, consistently showed two categories of stops in this condition. In the last two conditions, language set effects were further investigated by having a Thai speaker produce both English and Thai stops. The main objective was to see whether the accent of the Thai speaker has any effects on Thai participants' identification of English and Thai stops. Results show that Thai participants showed three and two categories of stops in their responses only when the Thai and the English stimuli that were produced by the Thai speaker were presented

in separate trials (Condition 4). The Thai participants were indecisive about using the Thai or the English patterns of stops when the Thai and English stimuli were interspersed within the same trial (Condition 5). These findings suggest that having Thai speakers produce both Thai and English stimuli affected Thai participants' identification of stops from these two languages. Although the results from Condition 4 indicate that Thai learners of English had a tendency to respond to English stops produced by a Thai speaker with two categories of voicing contrast, their responses were less consistent than when they were listening to English stimuli produced by native speakers of English.

Table 1. Responses to English and Thai [p] by Thai participants*

	English [p]		Thai [p]	
	/b/ judgments	/p/ judgments	/b/ judgments	/p/ judgments
Condition 1	48 (80%)	9 (15%)	0 (0%)	60 (100%)
Condition 2	37 (62%)	17 (29%)	1 (2%)	58 (97%)
Condition 3	31 (53%)	28 (47%)	28 (47%)	32 (53%)
Condition 4	39 (65%)	15 (25%)	-	-
Condition 5	30 (50%)	24 (40%)	4 (7%)	56 (93%)

* The heaviness of the line represents the degree of explicitness in English and Thai stop pattern in the judgments of Thai participants from most explicit to least explicit (most ambiguous).

6. Discussion

The performance of the Thai children and adults conforms to previous findings that inexperienced learners tend to perform according to their L1 inventory (Best 1995) while more experienced learners perform more like a native speaker of the target language (MacKain et al. 1981; Flege, Takagi & Mann 1996). In addition, the performance of the Thai children is consistent with the prediction made by the Perceptual Assimilation Model (PAM). This model states that L2 learners tend to rely on their L1 phonological inventory in their perception and production of non-native sounds (Best, McRoberts & Sithole 1988; Best 1995). Since the three phonetic categories of long-lead [b], short-lag [p] and long-lag [p^h] stops exist both in Thai and in English, this model predicts that Thai children will map English stop stimuli onto their counterparts in Thai. Thus, Thai children would show three categories of voicing contrast in their

responses, as found in Thai. Findings from the current study support this prediction.

Since the Perceptual Assimilation Model does not make any predictions regarding the performance of Thai adults who are experienced learners of English, their performance will be discussed in relation to the Speech Learning Model (SLM) (Flege 1992, 1995). This model attempts to account for second language speech learning across the life span. According to this model, the phonetic similarity between stops in Thai and English would lead Thai learners to assimilate the stop sounds of English to those of Thai by the so-called "equivalence classification" process. In contrast to the Thai children, the adult group exhibited only two categories of stops in their responses, as did the native speakers of English. Results have shown that they did not identify the English [b•] or [p] as voiceless unaspirated /p/. Rather, they identified them as voiced /b/. Hence, the performance by the Thai adult learners of English did not support the predictions made by the Speech Learning Model (SLM). The performance of the Thai adults contributes strong evidence that these learners know that there are two categories of stops in English. They showed polarization in their responses by not mapping the English [p] onto /p/, which is the nearest sound. Rather, they identified the English [p] as /b/.

There are some possible reasons for the polarization shown in the Thai adults' responses. First, the Thai adults might attempt to increase the perceived contrast between different stop consonants. When Thai speakers hear the English voiceless aspirated stop [p^h], they map it to the correct counterpart of this sound in Thai. However, they mapped both the voiced [b8] and plain stop [p] to the Thai voiced stop /b/. Although the VOT values of the English voiced and plain stimuli were more similar to those of the Thai plain stop /p/, Thai speakers did not choose this sound as their answer, but rather the Thai voiced stop /b/. The reason for this mapping might be that the difference in voice onset time between the Thai voiced and aspirated stops is greater than the difference in VOT value between the Thai aspirated and plain stops.

Another explanation for the observed polarization might be the effects of the teaching of Thai spelling of English words. English initial stops can be spelled in Thai as 'r' or '['. These two graphemes are pronounced as [p^h] and [b], respectively. Although Thai also has the plain stop /p/ in the language, this sound is not used in the Thai spelling of English initial stops. For example, an English word such as 'pin' would be spelled as 'rbo' (pronouced as [p^hɪn]). The initial letter is the sound [p^h]. A b-initial word like 'bin' will be spelled as '[bo', which is pronounced as [bɪn] in Thai. In this study, the English word-initial stimuli contained the three stops: [p^h], [b], and [p]. As described above, the target stimuli for the sound [p] were produced from words like 'sport.' Prior to the experiment, the initial [s] was removed. Thus, the participants heard the plain stop [p] as the initial consonant. It might be possible that Thai speakers were conscious of the system of Thai spelling of English words. Therefore, when they heard the English stops in initial position, they considered only two

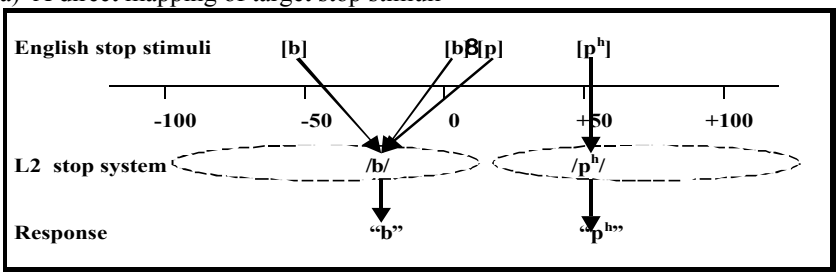
categories of stops, /p^{h}/ and /b/, in making their choices. Under this view, Thai speakers correctly know that English has only two categories of stops, but there is an unusual identification of the plain category as voiced, under the influence of the explicit instruction of the spelling system.

There is some anecdotal evidence that Thai speakers behave differently when they pronounce words from languages that are not explicitly taught to them, as in Thai speakers' pronunciation of Chinese proper names (the relevant Chinese languages lack fully voiced stops). For example, 'Beijing' [pe‡Ij&IN] and 'Tao' [tào] are pronounced [pa$kkI$$$N] and [ta‡o] respectively in Thai. Thai speakers do *not* show polarization in their pronunciation of the Chinese names. They pronounce the Chinese [p] and [t] as /p/ and /t/, respectively, not as voiced /b/ or /d/. In this case the borrowings conform to non-polarized systems. Of course, since this is only anecdotal evidence, further investigation is needed.

7. Mental Representation of Thai learners' English Stop System

This study has shown that the Thai adult speakers responded to the English stop stimuli similarly to the English speakers. The next question is whether these Thai learners have the same English stop system as that of the native speakers of English. Two possible answers are illustrated in Figure 4.

(a) A direct mapping of target stop stimuli



(b) An indirect mapping of target stop stimuli

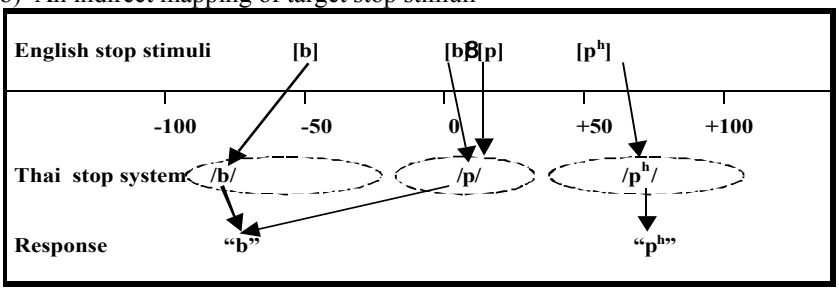


Figure 4. Direct and indirect mappings of English labial stops by Thai adults

The first possible answer (the direct mapping, shown in (a)) is that when Thai speakers learn English for some time, they are able to develop an L2 stop system that is the same as that of the English speakers (having two stop categories). Hence, when they hear the English stop stimuli, they directly map them onto the categories in this L2 system. They perceive both English voiced and voiceless unaspirated stops as the same sound /b/. Another possibility (the indirect mapping, shown in (b)) is that the Thai speakers develop an L2 stop system that is a copy of their native stop system (having three stop categories). When they hear the English stop stimuli, they first map them onto the three-category of stops in their system before relabelling them with the two categories for English. In the second case the Thai phonological system mediates between the stimuli and the L2 response. That is, Thai speakers would learn that [p] = [b] in English. They would still initially perceive a difference, but the difference would then be ignored or eliminated when the context language is English.

No evidence bearing on the choice of (a) vs. (b) in Figure 4 can be obtained from simple response data, as they agree at this level. There are several factors that might have led the Thai speakers to respond with two categories of stops, as did the English speakers, even if they perceived these English stimuli with three categories of stops. We cannot provide evidence to support either theory of the mapping of target stop stimuli to the L2 stop system from the behavioral data obtained in this study. We need instead to consider such evidence as brain-imaging data, which would allow us to examine the "sensory process" as distinct from the "decision process" (borrowing and adapting terminology from Signal Detection Theory (Green & Swets, 1966)). A project involving whole-head magnetoencephalographic (MEG) brain recordings is now being conducted (Idsardi & Imsri, 2002; Idsardi, Imsri & Phillips, in prep.).

8. Conclusions

The performance of the Thai children supports the Perceptual Assimilation Model, which posits that inexperienced learners tend to perceive non-native sounds according to the sound inventory of their L1. The performance of Thai adults, however, shows that they do not simply map English to Thai stops that are phonetically similar. Rather, they are aware of the language set of the stimuli, and are able to switch their responses according to the language they are listening to. These findings offer strong evidence that Thai speakers with more experience in English are aware of the difference between English and Thai stops. The Thai adults also show a polarization of their responses, which is compatible with several possible explanations.

The experiments reported here are, as always, limited by the methodology. Since only behavioral data was collected, only certain conclusions can be drawn, and we cannot reliably separate sensory and decision processes. We hope that further brain-imaging experiments will confirm and clarify these results and offer further insights into speech sound perception in first and second languages.

References

Best, C. T. (1995) A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues*. Timonium, MD: York Press.

Best, C. T. , McRoberts, G. & Sithole, N. (1988) The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perceptual Performance*, 14, 345-360.

Bohn, O. & Flege, J. E. (1993) Perceptual switching in Spanish/English bilinguals. *Journal of Phonetics*, 21, 267-290.

Caramazza, A., Yeni-Komshian, G., Zurif, E. & Carbone, E. (1973) The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America*, 54, 421-428.

Flege, J.E. (1995) Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues.* Timonium, MD: York Press.

Flege, J.E. & Eefting, W. (1987) Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication,* 6, 185-202.

Green, D.M. & Swets, J. A. (1966) *Signal detection theory and psychophysics*. New York: Wiley.

Idsardi, W. J. & Imsri, P. (2002) An MEG Study of Thai and English speakers' perception of stops: A preliminary report. Paper presented at Graduate Student Conference, University of Delaware.

Idsardi, W. J., Imsri, P. & Phillips, C. (in preparation) An MEG study of Thai and English speakers' perception of stops.

Imsri, P. (2002) Ph.D. dissertation, University of Delaware.

Kuijpers, C. T. L. (1996) Perception of the voicing contrast by Dutch children and Adults. *Journal of Phonetics,* 24, 367-382.

Lisker, L. & Abramson, A. S. (1964) A cross language study of voicing in initial stops: Acoustical Measurements. *Word*, 20, 384-422.

Lotz, J., Abramson, A. S., Gerstman, L., Ingemann, F. & Nemser, W. (1960) The perception of English stops by speakers of English, Spanish, Hungarian and Thai: A tape-cutting experiment. *Language and Speech*, 3, 71-77.

MacKain, K., Best, C., & Strange, W. (1981) Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 2, 369-390.

Strange, W., Akahane, R., Kubo, R., Trent, S. A., Nishi, K. & Jenkins, J. J. (1998) Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, 26, 311-344.

Werker, J. F. & Pegg, J. E. (1997) Adult and infant perception of two English phones. *Journal of the Acoustical Society of America*, 102, 708-721.

Williams, L. (1977) The perception of stop consonant voicing by Spanish-English bilinguals. *Perception and Psychophysics*, 21, 289-297.

Email: pimsri@udel.edu, idsardi@udel.edu