Appears in *Second Language Research* 14: 389-405(1998)

# Phonological transfer and levels of representation: The perceptual acquisition of Thai voice and aspiration by English and French speakers*

Suzanne Curtin, University of Southern California
Heather Goad, McGill University
Joseph V. Pater, University of Alberta

## Abstract

In this paper, we show that the generative phonological distinction between lexical and surface representation can explain apparently contradictory orders of acquisition of L2 voice and aspiration contrasts by native speakers of English. Cross-language speech perception research has shown that English speakers distinguish synthetic voice onset time counterparts of aspirated-unaspirated minimal pairs more readily than voiced-voiceless. Here, we present evidence that in the perceptual acquisition of the same Thai contrasts, English speakers acquire voicing before aspiration. These divergent orders are argued to be due to the levels of representation tapped by the methodologies employed in each case: surface representations in the earlier studies, and lexical in the present one. The resulting difference in outcomes is attributed to the presence of aspiration in surface, but not lexical representations in English (Chomsky and Halle, 1968). To address the further question of whether allophonic aspiration in English aids in the eventual acquisition of contrastive aspiration in Thai, we compare the developmental progression of the English learners to that of native speakers of French, whose L1 contains only a voicing contrast, and no surface aspiration. The performance of the anglophone group improves over time, suggesting that L1 surface features can be lexicalized in L2 acquisition, even though they are not initially transferred across levels.

## I Introduction

While the very existence of first language influences on second language syntax has been the source of some controversy (see Gass and Selinker, 1983; Odlin, 1989; and White, 1989 for discussion), no doubt has ever been expressed that the perception and production of second language speech is affected by the phonetics and phonology of a learner's first language (overviews of documented L1 effects in these domains can be found in Odlin, 1989; Leather and James, 1991; Major, 1994; and Strange, 1995). However, just as the existence of syntactic transfer does not imply that all aspects of the L1 are carried over to the L2 (Schwartz and Eubank, 1996), it remains to be determined what aspects of the phonetic and phonological systems of the L1 have an influence on the L2.

Several studies of production have suggested that only a particular subset of phonological rules are involved in transfer (Rubach, 1984; Singh and Ford, 1986; Dziubalska-Kołaczyk, 1986; Weinberger, 1994; Eckman and Iverson, 1997). Although the exact delineation of the set of transferable rules remains controversial, they are typically argued to be ones that apply at a late level in the L1 phonology (i.e., the post-cyclic or post-lexical rules in standard generative phonology; cf. Archibald, 1993 on transfer of lexical stress rules). There has also been some controversy over how best to characterize the influence of the L1 on L2 perception (see recently Hancin-Bhatt, 1994; Best, 1995; Flege, 1995; Rochet, 1995; and Brown, 1997), but in contrast with the production studies, the notion of phonological level has remained peripheral to the discussion. The present paper examines a topic that is related to both of these lines of earlier research, but which has thus far received little attention (cf. Brown, 1997): the role that differences between L1 lexical and surface representation play in L2 perceptual acquisition.

A lexical representation is the form of a word stored in long-term memory. In generative theory, it is standardly assumed that the phonological portion of lexical representations contains only distinctive, non-predictable features.[1] The textbook example of a difference between lexical and surface representation, and the one we will be focusing on in this study, is that of English aspiration. The distribution of aspiration in English is predictable, in that it occurs on all voiceless stops that begin a word-initial, or stressed syllable (1a). In other environments, voiceless stops are unaspirated (1b).

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1) | a. | [pʰæt] | 'pat' | [bæt] | 'bat' | [rəpʰíjt] | 'repeat' |
| | b. | [spæt] | 'spat' | [ræp] | 'rap' | [rǽpəd] | 'rapid' |

In Chomsky and Halle (1968) and following work, the predictability of aspiration is captured by representing all voiceless stops as unaspirated in the lexicon, and then invoking a rule to supply aspiration in the proper surface contexts.

| | | | | |
|---|---|---|---|---|
| 2) | Lexical Representation: | /pæt/ | /bæt/ | /ræp/ |
| | Aspiration Rule: | [pʰæt] | — | — |
| | Surface Representation: | [pʰæt] | [bæt] | [ræp] |

English, then, is treated as underlyingly identical to a language like French, which has a simple voiced-voiceless contrast:

| | | | |
|---|---|---|---|
| 3) | Lexical Representation: | /ba/ 'low' | /pa/ 'not' |
| | Surface Representation: | [ba] | [pa] |

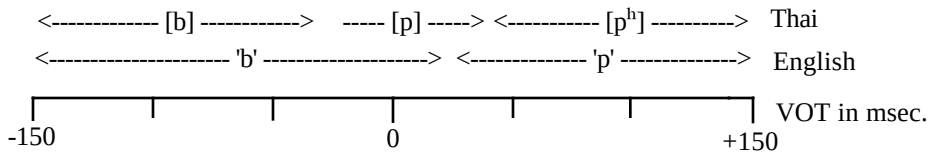
These languages differ only at the surface, in that French lacks an aspiration rule. Thai, on the other hand, is an example of a language that has both contrastive voicing and aspiration. Since we find triads of words that minimally differ in both voice and aspiration (4), neither of these features is predictable. In this language, both voice and aspiration are represented lexically.

| | | | | | | |
|---|---|---|---|---|---|---|
| 4) | [bèt] | 'fishhook' | [pèt] | 'duck' | [pʰèt] | 'spicy' |

The primary question we address in this study is whether English speakers first acquire contrastive voice or aspiration when learning Thai. If L2 lexical representations are initially constructed as in a learner's L1, then the voice contrast should be acquired first, since English lacks aspiration lexically. However, the cross-language study of voice onset time (VOT) perception suggests that these contrasts might be acquired in the opposite order.

Abramson and Lisker (1970) demonstrate that English speakers identify synthetic stimuli whose VOT values correspond to Thai unaspirated-aspirated minimal pairs as 'b' and 'p' respectively. English and Thai speakers were presented with synthetic stimuli whose voice onset time ranged from 150 msec. before release (-150), which represents the extreme end of the Thai voiced category, to 150 msec. after release (+150), at the edge of the aspirated category. The English speakers were asked to identify tokens as either 'b d g' or 'p t k'. Labials were consistently judged as 'b' up to +15 msec., and as 'p' from +35 on, with +25 marking the crossover point at which 'b' and 'p' judgements occurred with equal frequency. This is much closer to the Thai [p-pʰ] crossover of +40, than to the [b-p] crossover, which occurs in the prevoicing region, at -30 msec. Similarly, English 'd' and 't' judgements cross at +35 msec., closer to Thai [t-tʰ] (+45) than to [d-t] (-10). Figure 1 illustrates how English and Thai speakers divide up a labial VOT continuum based on their native language categories.

**Figure 1.** Identification of a VOT continuum by English and Thai speakers (based on Abramson and Lisker 1970)




This seminal finding has been replicated and extended in a number of experiments (see e.g. Strange, 1972; Pisoni et al., 1982). While Pisoni et al. (1982) do show that, with training, English speakers can distinguish VOT correlates of Thai voiced-voiceless minimal pairs, the first and more robust distinction is made at the aspirated-unaspirated boundary. These studies strongly support the conclusion that native speakers of English can perceive the aspiration distinction more easily than voicing, at least in terms of VOT. If the order of acquisition of lexical contrasts matches their relative perceptibility, then on the basis of this earlier research, we would expect aspiration to emerge first in native English speakers' L2 acquisition of Thai.

The present study departs from the VOT perception research in several ways. First, in order to examine the lexical acquisition of the Thai contrasts, we required subjects who had learned a lexicon of Thai words. To fulfil this need, we employed a training component in which monolingual subjects were taught a set of Thai words that included voice and aspiration 'minimal triads' like the one in (4) above. This not only allowed us to investigate the relatively unusual scenario of Thai being acquired as a second language in North America; it also enabled us to control for differences in the amount and context of L2 exposure. Second, in both training and testing, we used words produced by four Thai speakers, rather than stimuli generated by speech synthesis. This was done so as to simulate naturalistic acquisition as closely as possible within the experimental setting, and also to ensure that voicing and aspiration cues other than VOT would be included. Finally, we chose as our primary measure of subjects' discrimination abilities a task that taps lexical representations, as opposed to the phoneme identification and discrimination tasks used in the VOT research, which measure only the identification of sounds with native language categories, and the ability to distinguish minimally different sounds. Details of the training and testing procedures will be presented in the following section.

If, with this revised methodology, English speakers do not acquire the Thai aspiration contrast first, but instead, follow the order of acquisition derived from generative theory and first acquire voicing, then we can address a second question: Does the presence of aspiration in English L1 surface representations have any effect on speakers' ability to eventually lexically represent this feature? To investigate this issue, we included a comparison group of native speakers of French. French speakers would clearly be expected to acquire the Thai voice distinction before aspiration, since their L1 has a simple voiced-voiceless contrast at underlying and surface levels. If the presence of allophonic aspiration in English does play a role in the acquisition of contrastive aspiration in Thai, then there should be a difference between the English and French subjects in how quickly they acquire this Thai distinction.

## II Methodology

The subjects were nine monolingual speakers of Canadian English (mean age 25.5) and eight monolingual speakers of Canadian French as spoken in Québec (mean age 24). All reported having normal hearing, no previous exposure to Thai, and at least two years of post-secondary education. The tests were also completed by ten Thai native-speaker controls.

All of the stimulus words were CVC in shape. The contrast under investigation was restricted to onset position. Half of the stimuli were labial-initial and half began with coronals (velars were omitted because Thai lacks /ŋ/). Codas were restricted to /t k/ and /m n/. The vowels that were selected were maximally distinct from one another, /i a o/, and all were short. Half of the voice/aspiration triads were low-toned and half were mid-toned.

The words were paired with a picture of a noun for presentation to the subjects. (The original Thai sound-meaning pairings were not used, since many of the words were not nouns and could not be readily depicted.) Six triads produced a total of 18 words, as illustrated in (5).

5)

| | | | | | |
|---|---|---|---|---|---|
| bon | (sun) | pon | (watch) | pʰon | (shovel) |
| bàk | (pineapple) | pàk | (scissors) | pʰàk | (house) |
| bìt | (horse) | pìt | (cake) | pʰìt | (flower) |
| don | (boat) | ton | (tree) | tʰon | (fly) |
| dàk | (bucket) | tàk | (cat) | tʰàk | (shoe) |
| dam | (pear) | tam | (man) | tʰam | (car) |

The stimuli were elicited from four male native speakers of standard Bangkok Thai and were recorded on Digital Audio Tape at a 44.1 kHz sample rate, 16 bit resolution. Each word was placed in a carrier sentence to avoid a list effect; the stimulus words were randomized and interspersed with fillers. The recorded sentences were transferred to hard disk and the stimuli were spliced out using a digital audio editing program. For each word, three tokens produced by different speakers were selected that fell within the Thai VOT ranges provided in Lisker and Abramson (1964: 396).

Both training and testing were conducted on a Power Macintosh computer using PsyScope 1.1 (Cohen et al., 1993). The aural stimuli were presented at a comfortable listening volume over headphones, using the internal soundcard (44.1kHz/16 bits), and headphone amplifier of the Macintosh.

There were two days of training, separated and followed by testing days. The training and testing schedule is shown in (6).

6)

| Day 1 | Day 2 | Day 3 | Day 4 | Day 11 |
|---|---|---|---|---|
| Training | Review, Testing | Training | Testing | Review, Testing |

In each training session, the subjects were taught all 18 words. Training consisted of the presentation of the picture of an object on the computer screen at the same time that its name was presented aurally. To facilitate learning, the 18 words were presented in six groups of three; the words in each group were semantically-related but were completely different at the segmental level (e.g. [pʰàk] 'house' – [don] 'boat' – [tʰam] 'car'). Following the presentation of each set of three words, the subjects completed a word identification task that tested their knowledge of them. Subjects were shown two pictures on the computer screen, they heard one name, and had to press the key that corresponded to the appropriate picture. They were given positive feedback for correct responses. Through the course of the day's training, progressively larger sets of words were included in the word identification tasks. In the final identification task that included all 18 words, no subject's score for proportion correct fell below .94 on Day 1 nor below .98 on Day 3, which indicates that the training was successful in establishing sound-meaning pairings. To avoid possible influences of production difficulties on the establishment of perceptual categories, the subjects were never asked to pronounce the words. In addition, to avoid training the subjects on discrimination, none of the pairs of words that subjects were asked to choose between at this point were minimal

pairs. Discrimination training was omitted in the interest of examining how the contrasts would be acquired without explicit instruction (but see note 2). In this respect, the present study differs from training studies such as Strange (1972), Pisoni et al. (1982), Jamieson and Morosan (1986), and Rochet and Chen (1992), in which the training component focuses on the discrimination and identification of sounds, rather than on word learning.

Testing took place at the mid-point of training (Day 2), on the day immediately following training (Day 4), and after one week without any exposure to Thai (Day 11). The one-week gap was included to determine whether or not contrasts that had been acquired would be retained in long-term memory. The testing on Days 2 and 11 was preceded by a review of the words' meanings; this review was omitted on Day 4 because it was felt that the meanings would be firmly established by that point. In this paper, we report on the results of the two tests that were administered on all three of these days; the order in which they were administered was counterbalanced across subjects.[2]

To study the development of lexical representations, we employed a task in which subjects must choose between pictures of words that are in a minimal pair relationship, when presented with only a single word aurally (following Brown, 1997). To successfully complete this 'Minimal Pair' task, a subject must be able to perceptually distinguish the words, and must also have the contrast between the words stored in long-term memory. The pictures of the minimal pair are accompanied by a picture of a foil that differs phonetically in more than one segment from the other words. An example trial testing knowledge of the aspiration contrast would consist of the presentation of pictures for [tàk] (cat), [t$^{h}$àk] (shoe), and [bon] (sun) arrayed across the screen, along with the aural presentation of [tàk] (or [t$^{h}$àk]). Subjects respond by pressing a key that corresponds to the position of the appropriate picture on the screen.

There were 120 trials across five conditions, 24 each where the minimal pair was Asp-Plain, Plain-Vcd, or Asp-Vcd (where 'Asp' refers to unvoiced aspirated, 'Plain' to unvoiced unaspirated, and 'Vcd' to voiced unaspirated). There were also 24 'Place' controls which were included to test subjects' ability to perform the task with contrasts that are present in their native language (for example, subjects see pictures for [ton] (tree), [pon] (watch), [p$^{h}$àk] (house), and hear either [pon] or [ton]). For the final 24 trials, the correct response was the foil. Performance on the foil condition was used as a check on whether subjects were focused on the task, and as an additional means of determining whether they had learned the words. Trials were presented in random order.

The second set of results reported below comes from an ABX task in which a minimal pair 'AB' is presented aurally, followed by a third word 'X' that matches either A or B. The ABX task used here was designed to tap categorical phonemic judgements, as opposed to acoustic or phonetic discrimination. As in Gottfried (1983), and Flege, Munro, and Fox (1994), the tokens used for A, B and X were each produced by a different speaker, making a judgement based purely on acoustic similarity difficult. We also used an interstimulus interval between B and X of 750 msec., long enough to favour a phonemic analysis (Werker and Logan, 1985). There were 72 trials:
16 each of Asp-Plain, Plain-Vcd, and Asp-Vcd, and 24 Place controls.

## III Minimal Pair Results and Discussion

Since performance on the foil condition in the Minimal Pair task provides a measure of attention to task, we required a proportion of .90 correct for a subject to be included in the analysis. One anglophone was excluded on these grounds, reducing the number of English subjects from nine to eight. Performance on the foils for the remaining 16 subjects was as follows: on Day 2, it never fell below .91, and on Days 4 and 11, never below .95. As expected, performance on the place condition was also very good, with an overall mean proportion of correct responses of .90 for the English speakers, and .89 for the French. This indicates that the subjects were able to perform the task when the contrasts corresponded to native language categories.

Results for the conditions containing voice and aspiration contrasts are given in Table 1. For the Thai native speakers, the task was identical to that performed by the L2 learners except that orthographic representations were used in place of pictures. Orthography was used because many of the actual meanings for the words could not be easily depicted, and learning new meanings for existing lexical items would introduce a level of complexity to the task beyond that faced by the L2 learners. The Thai speakers performed the task almost perfectly, indicating that all of the stimuli could be reliably identified by native speakers (the few errors were due to a particular subject, rather than to any particular token).

| | **Asp–Plain** | | **Plain–Vcd** | | **Asp–Vcd** | |
|---|---|---|---|---|---|---|
| Time | English | French | English | French | English | French |
| Day 2 | .59 | .60 | .82 | .75 | .93 | .91 |
| Day 4 | .63 | .60 | .77 | .81 | .95 | .94 |
| Day 11 | .68 | .59 | .82 | .81 | .95 | .96 |
| | Thai .99 | | Thai .99 | | Thai 1.00 | |

**Table 1** Minimal Pair Task: Proportion Correct

The data for the non-native subjects provide a clear answer to our main research question: is voice or aspiration acquired first? A comparison of the results on Asp-Plain to those on Plain-Vcd reveals that while aspirated-unaspirated minimal pairs were discriminated by both groups at a level only slightly better than chance, performance on the voicing contrast was considerably better. A paired two-tailed t-test yields a significant difference between the results on these contrasts for both the French ($p<0.05$) and English subjects ($p<0.005$). Note that the very high scores on the Asp-Vcd pairs do not entail that the subjects are lexically contrasting both voice and aspiration. The more accurate discrimination of Asp-Vcd pairs than Plain-Vcd is likely due to the fact that the aspirates provide stronger cues to the voicing distinction than do the voiceless unaspirated segments. The Asp-Vcd results can be taken as indicating that the task can be performed reliably when the contrast does not pose any difficulty.

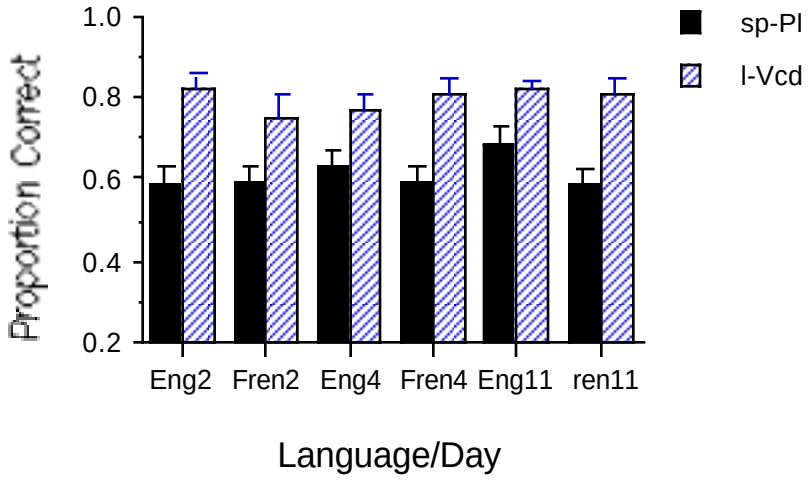
The difference between the results on the Asp-Plain and Plain-Vcd conditions is confirmed by an ANOVA examining the factors contrast (voice vs. aspiration), language, and testing day. There is an effect only for contrast ($F(1,14)= 7.538$, $p=0.001$), and no significant interactions between the factors. Further, Figure 2 also shows that the performance on the voicing contrast is consistently significantly higher for both the English and the French subjects; the scope of the error bars (which indicate Standard Error) never overlap.

An inspection of Figure 2 also shows that the English learners perform almost identically to their French counterparts, reflecting the lack of an effect for language as a factor in the ANOVA. With the exception of the Asp-Plain contrast on Day 11, the error bars overlap for the two groups' scores on each contrast at each testing time. Because accurate responses on the Minimal Pair task must be made on the basis of lexically stored representations, these data can be interpreted as showing that the voice contrast, rather than aspiration, is the contrast that both English and French speakers initially represent lexically when acquiring Thai. This lack of effect for language is as would be expected if L2 learners construct lexical representations based on the features that are contrastive in their L1, given that French and English both have a lexical voice contrast.

Since we have found that English speakers begin L2 Thai acquisition with a lexical voice contrast, we are in the position to address the question of whether the presence of surface aspiration has an effect on the eventual acquisition of an aspiration contrast. To do so, we now look at whether there

are differences between the development of the English and French subjects. If features can transfer across levels, then the English-speaking learners of Thai should acquire the lexical aspiration distinction more readily than the French speakers.

**Figure 2** Minimal Pair Task: Proportion Correct




As noted above, in an ANOVA including testing day, contrast, and language as factors, day does not reach significance as a main effect, nor are there significant interactions between day and other factors. However, Figure 2 does show the English subjects performing better on the aspiration contrast than the French speakers on the last day of testing, with no overlap between the scope of the error bars. Somewhat surprisingly, the English speakers only showed an improvement at Day 11, one week after the training occurred, and not at Day 4, immediately after the training. This could be due to the fact that on Days 2 and 11, but not on Day 4, the testing was immediately preceded by a review of the words' meanings (cf. (6) above). To further investigate whether there is a developmental difference between the groups that is not being captured by the ANOVA, we measured development by subtracting subjects' Day 2 scores from their Day 11 scores. The resulting scores were then subjected to a Kolmogorov-Smirnov test, which like a chi-square is a goodness-of-fit test, but is better suited to cases when the hypothetical distribution is continuous or the sample size small, both of which are true here. This test yielded a significant difference between English and French speakers ($p<0.002$).

One might wonder if this improvement on the part of the English speakers is a general trend, or is displayed by only a few individuals. It appears that the latter is the case; three learners showed an improvement of greater than 15 points (mean +.24, range +.17 to +.33) while the other five improved by 8 points or less (mean: .00, range -.08 to +.08). A Kolmogorov-Smirnov test shows that these two sets of subjects are significantly different ($p<0.001$).

Although the small sample size and short duration of this study preclude any firm conclusions, the developmental difference between some of the English and French subjects suggests that the presence of surface aspiration in English may facilitate the establishment of a lexical aspiration contrast in the L2 acquisition of Thai. It appears, then, that L1 surface features can be lexicalized in L2 acquisition.

While a subset of English speakers did show improvement on the aspiration contrast, the relatively poor performance of the group as a whole, especially at the outset, is unexpected, given the earlier research reviewed in the Introduction which found better discrimination performance on VOT correlates of the aspiration contrast than of voice. We have suggested that the reason that English speakers perform better on the voicing contrast in the Minimal Pair task is that, in constructing L2 lexical representations, they are initially making use of only those features that are present lexically in their L1. Since the identification and discrimination tasks of cross-language VOT perception research do not require lexical access, and in fact cannot require it, as no meaning is associated with the stimuli, it is not surprising that they would be performed on the basis of surface phonetic similarity. As aspiration is present only in surface representations, the generative lexical-surface distinction allows us to reconcile these apparently contradictory results.

As the present study also differs from the earlier VOT research in that it employs natural rather than synthetic stimuli, it might be argued that the results obtained in our Minimal Pair task are due to purely phonetic factors, rather than being due to the underlying absence of aspiration in English. Perhaps the voice distinction in Thai is simply easier to perceive than aspiration, due to the influence of cues other than VOT that are present in the naturalistic stimuli. We cannot address this alternative directly, as we included no measure of pure acoustic or phonetic discrimination in our testing phase; in fact, it is difficult to construct such a measure using natural stimuli produced by multiple speakers. Nevertheless, our ABX results do shed light on this issue.

## IV ABX Results and Discussion

As in the Minimal Pair task, we will not present the Place results for the ABX task in detail, but as expected, the scores were high: .92 and .94 correct for the English and French subjects respectively. Results on the voice/aspiration triads are given in Table 2. The Thai native speakers' average performance on all three conditions is very high, though interestingly, not quite as high as for the Minimal Pair task.

| | **Asp–Plain** | | **Plain–Vcd** | | **Asp–Vcd** | |
|---|---|---|---|---|---|---|
| Time | English | French | English | French | English | French |
| Day 2 | .84 | .64 | .83 | .78 | .99 | .96 |
| Day 4 | .77 | .67 | .73 | .77 | .99 | .98 |
| Day 11 | .79 | .59 | .70 | .88 | .88 | .98 |
| | Thai .92 | | Thai .91 | | Thai .96 | |

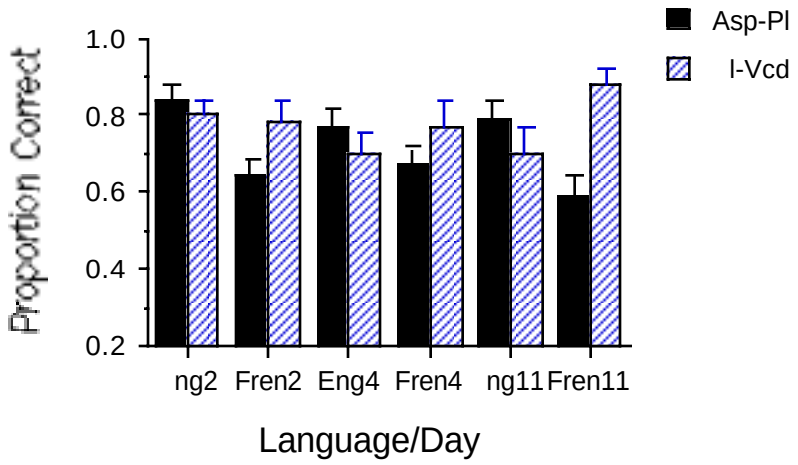
**Table 2** ABX Task: Proportion Correct

For both the English and French speakers, we again find extremely good performance on the easily discriminable combination of voicing and aspiration contrasts (Asp-Vcd). However, in the crucial Asp-Plain and Plain-Vcd conditions, we see differences between the language groups.[3] The French speakers perform better on the voice contrast than on aspiration, as they did in the Minimal Pair task. The English speakers, on the other hand, perform similarly on voice and aspiration in the ABX task, whereas in the Minimal Pair, their performance on aspiration was significantly worse than on voice. For the ABX task, a paired two-tailed t-test on contrast yields significance in the French group ($p<0.05$), but not in the English (recall that in the Minimal Pair task, significance was reached for both groups). In an ANOVA, language, contrast, and testing day all fail to reach significance as independent factors, but language and contrast do interact ($F(1,14)=13.025$, $p=0.003$). The interaction observed between language and contrast in the ABX task and the lack of any such

interaction in the Minimal Pair task is exactly as expected under a level-based account, since the phonologies of English and French differ in the presence of aspiration at the surface.

The performance of the English and French groups is illustrated in Figure 3. For the English subjects, the scope of the error bars for the two contrasts overlaps, confirming the lack of a significant effect for contrast, while for the French subjects, the performance on the voice contrast remains significantly higher at all testing times.

**Figure 3** ABX Task: Proportion Correct




In showing equivalent performance on voice and aspiration, the results for the English speakers on the ABX task fall somewhere between those of the prior VOT research and those of our Minimal Pair task. Our interpretation of these intermediate results is that subjects performing this ABX task sometimes rely on lexical representations, and sometimes on surface ones. As discussed in the methodology section, our ABX task was designed to tap what is referred to in the literature as phonemic perception; the goal was to make the task difficult to perform on the basis of surface similarity by using tokens produced by multiple speakers and a long interstimulus interval. However, in contrast with the Minimal Pair task, the ABX task metholodogy does not *require* lexical access. As subjects are presented with aural stimuli only, they can make a decision on the basis of phonetic similarity rather than relying on stored representations.

We do not deny that the acoustic make-up of the naturalistic Thai stimuli used in the present study could be a contributing factor for the divergence of our Minimal Pair results from those of the earlier VOT research. It seems unlikely that it is the sole factor, however, given the differences in outcome between the present ABX task and the Minimal Pair task, which used the same set of naturalistic stimuli, and were completed by the same subjects.

## V Conclusions

In this study of the perceptual acquisition of Thai voice and aspiration contrasts, we have found evidence that English and French learners both lexically represent voice contrasts before aspiration, in that they perform much better on a Minimal Pair identification task when faced with a choice between a pair of words differing only in voice, than when the words differ only in aspiration. This result was contrasted with the findings of cross-language VOT perception research, which has

shown that English speakers distinguish VOT correlates of the aspiration distinction far better than voicing. The divergence between these results was argued to be due to differences in the level of representation tapped by the tasks employed in each case; the Minimal Pair task accesses lexical representations that lack aspiration, while the phoneme identification and discrimination tasks of the VOT research access surface representations that contain aspiration. This account was supported by results from an ABX discrimination task that used the same stimuli as the Minimal Pair task, which yielded better aspiration discrimination by the English subjects, but not by the French subjects whose L1 phonology lacks aspiration entirely. This set of results leads to the conclusion that L2 learners initially construct lexical representations that make use of only those features that are present lexically in the L1, even though they may be able to discriminate other L2 contrasts on the basis of surface features, and may eventually lexicalize these surface features. In a recent study of L1 perceptual acquisition, Werker and Stager (1996) also find what seems to be a difference between what learners can perceive, and what they lexically encode. This shared finding of a distinction between lexical and non-lexical tasks in developmental speech perception research opens up numerous avenues for further research.

**References**


Abramson, A. and Lisker, L. 1970: Discriminability along the voicing continuum: Cross-language tests. In Hala, B., Romportl, M. and Janota, P., editors, *Proceedings of the Sixth International Congress of Phonetic Sciences.* Prague: Academia, 569-573.

Archibald, J. 1993: *Language learnability and L2 phonology: The acquisition of metrical parameters.* Dordrecht: Kluwer.

Best, C. 1995: A direct realist view of cross-language speech perception. In Strange, W., editor, *Speech perception and linguistic experience: Issues in cross-language research,* Timonium, Maryland: York Press, 171-204.

Brown, C. 1997: The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research* (in press).

Caramazza, A. and Yeni-Komshian, G. 1974: Voice onset time in two French dialects. *Journal of Phonetics* 2, 239-45.

Caramazza, A., Yeni-Komshian, G., Zurif, E. and 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.

Chomsky, N. and Halle, M. 1968: *The sound pattern of English.* New York: Harper and Row.

Cohen, J.D., MacWhinney, B., Flatt, M. and Provost, J. 1993: PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments and Computers,* 25(2), 257-271.

Dziubalska-Kołaczyk, L. 1986: Phonological rule typology and second language acquisition. In James, A. and Leather, J., editors, *Sound patterns in second language acquisition,* Dordrecht: Foris, 193-205.

Eckman, F. and Iverson, G. 1997: Structure preservation in interlanguage phonology. In Hannahs, S.J. and Young-Scholten, M., editors, *Focus on phonological acquisition,* Amsterdam: John Benjamins, 183-207.

Flege, J.E. 1995: Second language speech learning: Theory, findings, and problems. In Strange, W., editor, *Speech perception and linguistic experience: Issues in cross-language research,* Timonium, Maryland: York Press, 233-277.

Flege, J.E., Munro, M. and Fox, R. 1994: Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America* 95, 3623-3641.

Gass, S. and Selinker, L., editors. 1983: *Language transfer in language learning.* Rowley, MA: Newbury House.

Gottfried, T. 1983: Effects of consonant context on the perception of French vowels. *Journal of Phonetics* 12, 91-114.

Hancin-Bhatt, B. 1994: Segment transfer: A consequence of a dynamic system. *Second Language Research* 10, 241-269.

Idsardi, W. 1997: Against lexicon optimization. Talk delivered at the Montreal-Ottawa-Toronto Phonology Workshop, University of Toronto.

Inkelas, S. 1994: The consequences of optimization for underspecification. Ms., University of California, Berkeley.

Itô, J., Mester, R.A. and Padgett, J. 1995: Licensing and underspecification in Optimality Theory. *Linguistic Inquiry* 26, 571-613.

Jamieson, D. and Morosan, D. 1986: Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology* 43, 88-96.

Leather, J. and James, A. 1991: The acquisition of second language speech. *Studies in Second Language Acquisition* 13, 305-341.

Lisker, L. and Abramson, A. 1964: A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20, 384-422.

Major, R. 1994: Current trends in interlanguage phonology. In Yavaş, M., editor, *First and second language phonology,* San Diego: Singular Publishing, 181-204.

Munro, M. 1987: Voicing contrasts in French and English labial stops. Master's thesis, University of Alberta.

Odlin, T. 1989: *Language transfer: Cross-linguistic influence in language learning.* Cambridge: Cambridge University Press.

Pisoni, D., Aslin, R., Perey, A. and Hennessy, B. 1982: Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance* 8, 297-314.

Prince, A. and Smolensky, P. 1993: Optimality Theory: Constraint interaction in generative grammar. Ms., Rutgers University and University of Colorado.

Rochet, B. 1995: Perception and production of second-language speech sounds by adults. In Strange, W., editor, *Speech perception and linguistic experience: Issues in cross-language research,* Timonium, Maryland: York Press, 379-410.

Rochet, B. and Chen, F. 1992: Acquisition of French VOT contrasts by native speakers of Mandarin Chinese. In Ohala, J. et al., editors, *ICSLP 92 (International Conference on Spoken Language Processing) Proceedings,* Edmonton: University of Alberta, 273-276.

Rubach, J. 1984: Rule typology and phonological interference. In Eliasson, S., editor, *Theoretical issues in contrastive phonology,* Heidelberg, Julius Groos Verlag, 37-50.

Schwartz, B. and Eubank, L. 1996: What is the 'L2 initial state'? *Second Language Research* 12, 1-5.

Singh, R. and Ford, A. 1986: Interphonology and phonological theory. In James, A. and Leather, J., editors, *Sound patterns in second language acquisition,* Dordrecht: Foris, 163-172.

Strange, W. 1972: The effects of training on the perception of synthetic speech sounds: Voice onset time. Doctoral dissertation, University of Minnesota.

Strange, W., editor, 1995: *Speech perception and linguistic experience: Issues in cross-language research.* Timonium, Maryland: York Press.

Weinberger, S. 1994: Please speak faster, I can't understand you: On the L2 transfer of fast speech rules. Talk delivered at Second Language Research Forum, Concordia University and McGill University, Montreal.

Werker, J. and Logan, J. 1985: Cross-language evidence for three factors in speech perception. *Perception and Psychophysics* 37, 35-44.

Werker, J. and Stager, C. 1996: Developmental changes in infant speech perception and early word learning: Is there a link? To appear in Pierrehumbert, J. and Broe, M., editors, *Papers in Laboratory Phonology V.* Cambridge: Cambridge University Press.

White, L. 1989: *Universal grammar and second language acquisition.* Amsterdam: John Benjamins.

* Authors' names are in alphabetical order. We would like to particularly acknowledge C. Brown for her collaboration in the initial stages of this research project. We would also like to thank our Thai consultants, especially V. Siripark, for help with construction of the stimuli; J. Kurtz for help with the recording; L. Sabourin for help with PsyScope; A. Bahatnagar, V. Lapointe, N. Stone, P. Royle and especially L. Vinette and S. Zyglidopolous for help with subject recruitment; M. Joanisse for much help with the stats, as well as J. Werker for supplying the facilities for testing the Thai subjects. For useful discussion and comments, we would like to acknowledge T. Nearey, G.L. Piggott, D. Pisoni, D. Pulleyblank, J. Werker, L. White, and one anonymous reviewer. This work was supported by an FCAR grant to H. Goad, by SSHRC grants to G. Jarema and L. White, and by a SSHRC post-doctoral fellowship to J. Pater. E-mail addresses: curtin@scf.usc.edu, goad@leacock.lan.mcgill.ca, joe.pater@ualberta.ca.

[1] It should be pointed out, however, that this assumption has recently been challenged within the generative literature. Specifically, work within constraint-based approaches such as Optimality Theory has opened up the possibility of specifying predictable features underlyingly (see Prince and Smolensky, 1993; Inkelas, 1994; Itô, Mester and Padgett, 1995; cf. Idsardi, 1997).

[2] There were several other tasks (ABX and 4IAX) that were administered during these testing phases, and in a pretest. However, these tasks are not directly comparable to the Minimal Pair and ABX tasks that we discuss in this paper because they used a different set of stimuli (18 different words), and because they were not given on the same three testing days. Therefore, for reasons of space, we do not report on those results here. However, it is worth noting that the use of discrimination tasks in the pretest may have yielded some training effects.

[3] Interestingly, Caramazza and Yeni-Komshian (1974) also found differences in the perception of the VOT continuum between speakers from the same two language groups, Canadian English and Canadian French. Specifically, the Canadian French VOT boundary, averaged for labials, alveolars, and velars, is about 20 msec. to the left of the Canadian English boundary. The crossover boundary is not as sharp for the Canadian French speakers as it is for the Canadian English or European French speakers. Caramazza and Yeni-Komshian (1974) and Caramazza, Yeni-Komshian, Zurif and Carbone (1973) take this to indicate that VOT is no longer "phonemically functional" in Canadian French. It is clear that VOT is not as strong a cue for Canadian French speakers as it is for the other groups, but as Munro (1987: 6) argues, it is likely that the case against VOT is overstated, since it is only over a relatively small range of stimuli that VOT does not cue consistent judgements.