

*Slightly revised version to appear in Second Language Research 19;3, 2003*

## **The Perceptual Acquisition of Thai Phonology by English Speakers: Task and Stimulus Effects\***

Joe Pater, *University Of Massachusetts*

### **Abstract**

This paper presents a follow-up to Curtin *et. al's* study of the perceptual acquisition of Thai laryngeal contrasts by native speakers of English, which found that subjects performed better on contrasts in voice than aspiration. This finding, surprising in light of earlier cross-linguistic VOT research, was attributed to the fact that the task tapped lexical representations, which are unspecified for aspiration, according to standard assumptions in generative phonology. The present study further investigated possible task effects by examining the discrimination and categorization of the same stimuli in various experimental conditions. Stimulus effects were also investigated by performing token-based analyses of the results, and by comparing them to acoustic properties of the tokens. The outcome of the discrimination experiment was the opposite of the earlier study, with significantly better performance on contrasts in aspiration than voice, even on a lexical task. A second finding of this experiment is that place of articulation interacts with the perception of the laryngeal distinctions; the aspiration distinction is discriminated better on the labials, and voice on alveolars. A parallel effect of place of articulation was also found in a categorization experiment.

### **1. Introduction**

Curtin, Goad, and Pater (1998) studied the initial stages of English speakers' acquisition of the Thai three-way stop voicing distinction (e.g. /b-p-p<sup>h</sup>/) by teaching monolingual subjects a set of 18 Thai words, and by having them perform several tasks that tested their ability to distinguish minimal pairs. In a picture selection task adapted from Brown (1998), subjects chose between pictures on the basis of an aurally presented word. When English speakers were faced with a choice between pictures that corresponded to minimal pairs differing in the feature [+/-voice] (i.e. voicing lead: /b-p/ or /d-t/), the mean proportion of correct responses was .80, averaged across testing times. When the members of the minimal pair differed in [+/- aspiration] (i.e. spread glottis or voicing lag: /p-p<sup>h</sup>/ or /t-t<sup>h</sup>/), the subjects scored .64 correct overall, a difference that was statistically significant.

That the subjects performed better on the voice contrast than aspiration is surprising in light of earlier research on the perception of voice onset time (VOT) using synthetic stimuli. This research has consistently shown that English speakers label synthetic VOT stimuli as 'b' in the

---

\* I would especially like to thank Wolf Wikeley for assistance in testing and phonetic analysis, and John Kingston and Terrance Nearey for discussion and help with statistics. Thanks also to Lesley Ewen, Kevin Hynna and Tina Thomas for their assistance in testing, to Mary Daly for administrative help, to Suzanne Curtin, Heather Goad, and audiences at ICPHS 1999, MIT, and the University of Massachusetts, Amherst, for helpful discussion, and to the anonymous reviewers for their helpful comments. This research was supported by SSHRC postdoctoral fellowship 756-96-0521 and SSHRC research grant 410-98-1595.



Thus, the Curtin *et al.* results may be seen as supporting the position that predictable features are absent from lexical representations (see also Lahiri and Marslen-Wilson 1992; cf. Ohala and Ohala 1995). However, if the underspecification of aspiration in lexical representations impedes English speakers' perception of Thai aspiration, then one might wonder how they are able to map the English surface distinction to underlying voicing, when the English surface distinction is closer to the Thai aspiration contrast. This issue is pursued further in section 4.

As opposed to rule-based phonology, there is no inherent connection between the predictability of a feature and its presence in the lexicon in Optimality Theory (OT), a constraint based theory of generative phonology (Prince and Smolensky 1993). In fact, under one proposal within OT, termed Lexicon Optimization (Prince and Smolensky 1993; Itô, Mester, and Padgett 1995), predictable features end up specified in lexical representations. To illustrate both the independence of predictability from specification in OT and Lexicon Optimization, I will sketch an OT account of the distribution of aspiration in English. For further discussion of the treatment of allophony in OT, and a general introduction to the theory, see Kager (1999) and McCarthy (2002).

Let us first assume a context-free markedness constraint \*ASP that is violated by any occurrence of the feature [+Aspiration]. If this constraint were undominated, aspiration would be ruled out categorically in the language (as in most Romance languages, for instance). To account for allophonic aspiration this constraint must be outranked by a contextual markedness constraint that demands aspiration in particular environments; in English, on voiceless stops in initial position of stressed and word-initial syllables. We can refer to this fortition constraint, which “strengthens” voiceless consonants by aspirating them, as FORT. In addition, both of these constraints must dominate the faithfulness constraint that requires underlying and surface identity of aspiration specification, IDENT(ASP); otherwise aspiration would be contrastive. The following pair of tableaux shows how the dominance of FORT results in word-initial voiceless stops being aspirated, no matter what the underlying specification of aspiration is:

( 3 )

/peɪ/	FORT	*ASP	IDENT(ASP)
☞ [p <sup>h</sup> eɪ]		*	*
[peɪ]	*!		

/p <sup>h</sup> eɪ/	FORT	*ASP	IDENT(ASP)
☞ [p <sup>h</sup> eɪ]		*	
[peɪ]	*!		*

In positions not targeted by the FORT constraint (i.e. positions other than word-initial, or stressed syllable-initial), segments emerge as unaspirated irrespective of their underlying specification, due to the ranking of \*ASP over IDENT(ASP):

( 4 )

/eɪp/	FORT	*ASP	IDENT(ASP)
[eɪp <sup>h</sup> ]		*	*
☞ [eɪp]			

/eɪp <sup>h</sup> /	FORT	*ASP	IDENT(ASP)
[eɪp <sup>h</sup> ]		*	
☞ [eɪp]			*

An OT grammar serves as a filter, in that it maps any input lexical representation to an output surface representation that is well-formed in the language. Markedness constraints hold of the output level only, thus avoiding the *duplication problem* of theories such as standard generative

phonology, which posit similar constraints or rules at both the surface and underlying levels (Kenstowicz and Kisseberth 1977).

In the theory elaborated so far, a learner of English might adopt as the lexical representation for an initial voiceless stop either the aspirated or the unaspirated form. While such indeterminacy is not necessarily problematic (it has no obvious empirical consequences), Prince and Smolensky (1993) do propose a resolution of it, termed Lexicon Optimization. Under this proposal, for any set of inputs that map to the same output, the one that best satisfies the constraint hierarchy will be chosen as the actual lexical form. Itô, Mester, and Padgett (1995) introduce the *tableau de tableaux* to illustrate how Lexicon Optimization works. This tableau takes all input-output mappings that converge on the same output, and shows the constraint violations incurred by each mapping. For our example of initial voiceless stops in English, the following tableau shows that Lexicon Optimization will choose the aspirated lexical form, as this one will result in no violation of the faithfulness constraint IDENT(ASP).

(5)

Input	Output	FORT	*ASP	IDENT(ASP)
/pei/	[p <sup>h</sup> ei]		*	* !
☞ /p <sup>h</sup> ei/	[p <sup>h</sup> ei]		*	

In general, Lexicon Optimization winds up producing lexical forms that match surface forms, since this avoids violations of faithfulness constraints. As in the aspiration example, the result is that both predictable and non-predictable information is encoded lexically.

Inasmuch as the Curtin *et al.* findings support the traditional view that the lexicon is free of phonologically predictable information, they also seem to provide evidence against Lexicon Optimization (see also Idsardi 1997; though cf. the discussion in section 4 below). It is therefore of theoretical interest whether these experimental results are robust. The first experiment presented in this paper aimed to replicate that study, using the same stimuli, but with slightly different tasks. The “lexical” and “non-lexical” tasks were made more similar to one another, so that they differed only in whether lexical access was required. Both were of the form XAB, where three stimuli are presented sequentially, and the subject is asked to determine whether the second or the third matches the first. The lexical tasks contained a picture that had to be matched to its aural label, while the non-lexical task required only a comparison between three aural stimuli.

Beyond the specific issue of whether predictable information is stored in the lexicon, the studies undertaken here were more broadly aimed at further investigating the relative contributions of task and stimulus properties to the divergence between the Curtin *et al.* results and those of prior VOT research. In support of this aim, a categorization task of the type commonly employed in cross-linguistic VOT studies was run using the stimuli from the Curtin *et al.* study.

In the results to be presented below, we will see that in both the “lexical” and “non-lexical” XAB tasks, the aspiration contrast was discriminated better than the voicing contrast, a result that is in line with those from experiments using synthetic VOT continua, and is counter to the findings of Curtin *et al.* While this might be seen as supporting lexicon optimization, in the concluding discussion I suggest that this type of experiment may not in fact be suitable for testing claims about the specification of lexical representations. A further result of the XAB tasks is an interaction between place and voicing, such that voice is discriminated better in alveolars

than in labials, and aspiration is discriminated better in labials than in alveolars. This is mirrored in the categorization results, in that the Thai voiceless unaspirated alveolars were more often labeled by the English listeners as ‘t’ than ‘d’, while the voiceless unaspirated labials were more often labeled as ‘b’ than ‘p’. cannot be explained by VOT differences, suggesting that other cues play a role in the discrimination and categorization of voicing contrasts.

## 2. Lexical and non-Lexical Discrimination Tasks

### 2.1. Method

The subjects were 10 native speakers of Canadian English, undergraduate university students with no exposure to a second language beyond high school French or German, and no linguistics training. They were recruited through response to posted advertisements, and were paid \$20 each for their participation.

The aural stimuli consisted of a set of 18 Thai words recorded from 4 speakers of standard Bangkok Thai; recording details are provided in Curtin *et al.* Three tokens, each from a different speaker, were selected for each word for the Curtin *et al.* study, and were also used here. The 18 words were made up of 6 sets of 3 monosyllabic words differing only in the voicing of the initial consonant. Half of the initial consonants were alveolars and half labials (Thai has no voiced velar), and half of the words had low tone and half mid.

The aural stimuli were paired with pictures of objects for presentation to the subjects (these pictures did not depict the original Thai meanings, since these were often not picturable). Representative sound-meaning pairings are provided in (6).

- (6) a. bít (flower)                      pít (car)                      p<sup>h</sup>ít (dog)  
    b. dam (hat)                         tam (horse)                    t<sup>h</sup>am (train)

Training and testing took place over two consecutive days, lasting about one hour a day. All stimulus presentation and data collection was performed using PsyScope 1.2 (Cohen et al. 1993) on a Macintosh computer. Aural stimuli were converted from 44.1 kHz/16 bit digital format to analog using an external Lucid ADA 1000 converter, were amplified using a Rotel RA-970BX amplifier, and were presented to the subject at a comfortable listening volume over Sony DR-S3 headphones. Visual stimuli were presented on a 17" color monitor about two feet away from the subject.

In the first day, subjects were taught the meanings of the words through repeated exposure to sound-picture pairs. The training was interspersed with tests in which subjects had to indicate if a presented sound-meaning pairing was correct. They were given immediate feedback on the correctness of the answer, and the sound matching the picture was repeated after each of these test trials, regardless of whether the subjects had answered correctly or not. The amount of exposure was held constant across subjects and stimuli, and minimal pairs were kept as far apart as possible during training.

Before the actual testing phase began on the second day, subjects performed a shortened version of the training procedure as a review. Three discrimination tasks were then administered in random order, followed by a categorization task to be described in section 3 below. The discrimination tasks were all of the form XAB, where the subject has to decide whether 'X' matches either 'A' or 'B'. Tasks varied only in whether the stimuli making up 'X', 'A', and 'B' were sounds or pictures. In all cases the stimuli were presented sequentially, with 1500 ms. from the

beginning of X to the beginning of A, and 1250 ms. from the beginning of A to the beginning of B. Pictures were presented for 1000 ms. The combinations of sounds and pictures used in each task are given in (7), along with the task names.

(7) *XAB Tasks*

- Sound Sound Sound (SSS)
- Sound Picture Picture (SPP)
- PictureSound Sound (PSS)

The tasks with pictures (SPP and PSS) require lexical access for successful completion, and are in this way parallel to the picture selection task used in Curtin *et al.* Their picture selection task differed, however, in that the aural and visual stimuli were presented simultaneously. As such, the task demands of the SPP and PSS are closer to the SSS task than the picture selection task would be. In each task, there were 72 trials in which A and B were members of a minimal pair, plus 54 foils in which A and B were not in a minimal pair relationship. The foils were included not only as distractors, but also as a measure of task difficulty, and of knowledge of the sound-meaning pairings. All 126 trials were presented in random order.

Examples of trials containing minimal pairs differing in voice and aspiration as well as a foil trial, are provided in (9). Subjects were not tested on discrimination of pairs differing in both voice and aspiration (e.g. [b] vs. [p<sup>h</sup>]), since these were found to present no difficulty in Curtin *et al.*'s study.

- (9)
- |         |      |      |              |
|---------|------|------|--------------|
| X       | A    | B    |              |
| a. b̩t  | b̩t  | p̩t  | (Voice)      |
| b. p̩t  | p̩t  | b̩t  | (Voice)      |
| c. ph̩t | p̩t  | ph̩t | (Aspiration) |
| d. p̩t  | ph̩t | p̩t  | (Aspiration) |
| e. ph̩t | dam  | ph̩t | (Foil)       |

The location of the match for X was balanced between A and B within all trial types. For tasks involving sounds, each of the words was spoken by a different speaker. Trials were each presented once in random order.

2.2 *Results*

Table 1 presents the group mean proportion correct and standard deviations for the voice contrast trials (VCE), the aspiration contrast trials (ASP), and the foils (FOIL) in each of the three tasks.

	SSS	SPP	PSS
VCE	0.71 (0.12)	0.53 (0.11)	0.72 (0.12)
ASP	0.84 (0.12)	0.52 (0.09)	0.83 (0.10)
FOIL	0.97 (0.04)	0.98 (0.04)	0.98 (0.02)

Table 1. Overall results

The high scores on the foils indicate that the subjects had learned the meanings of the words, and were capable of performing all of the tasks. However, when the pictures in the SPP task represented words in a minimal pair relationship, subjects performed only at chance. This occurred consistently across subjects, and across all types of tokens. It would seem that this task presented a higher degree of difficulty than PSS. Since the subjects performed at chance in all of the conditions in the SPP task, the results for this task will be left out of subsequent analyses.

In the remaining tasks, we find that subjects performed significantly better on the aspiration distinction than on voice, as illustrated in the following graph. Error bars indicate a 95% confidence interval; there is very little overlap between the confidence intervals for voice and aspiration in either of the tasks:

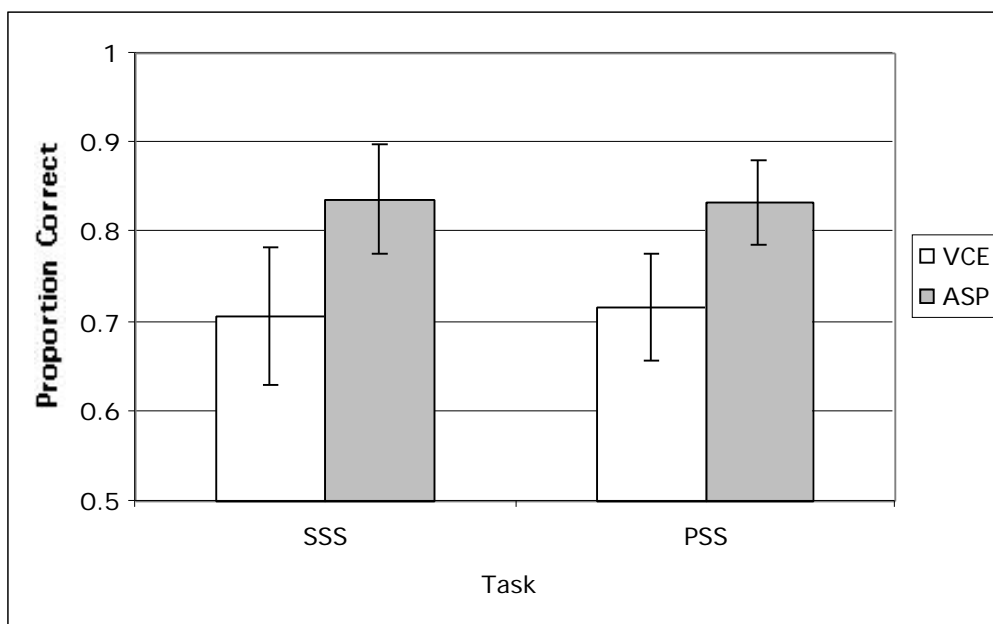


Figure 2: Results for voiced vs. aspirated stops

An analysis of variance was performed with three within-subject factors: voicing contrast (voice vs. aspiration), place of articulation (alveolar vs. labial), and task (SSS vs. PSS). The analysis showed a significant main effect only for voicing contrast ( $F_1(1,9) = 8.34, p < .02$ , by subject;  $F_2(1,17) = 10.9, p = .004$ , by item). There was also a significant two-way place by contrast interaction ( $F_1(1,9) = 19.13, p = .002$  by subject;  $F_2(1,17) = 59.38, p < .001$ , by item). No other even marginally significant main effects or interactions were observed. Thus, in both the task that requires lexical access (PSS) and the one that does not (SSS), aspiration is discriminated better than voice.

The place by contrast interaction detected by the ANOVA is evident in the means in Table 2, which separates the trials according to place of articulation. Within the alveolars, there is little difference between the voice and aspiration contrasts, whereas there is an especially marked difference between discrimination of voice and aspiration contrasts on labials.

	ALVEOLAR		LABIAL	
	SSS	PSS	SSS	PSS
VCE	0.78 (0.15)	0.81 (0.13)	0.63 (0.11)	0.62 (0.13)
ASP	0.78 (0.14)	0.77 (0.15)	0.9 (0.13)	0.9 (0.08)

Table 2. Results for SSS and PSS tasks according to place

Considering the results within each type of voicing contrast, one finds place-related differences. Discrimination of the voice contrast is better in alveolars than labials, while discrimination of the aspiration contrast is better in labials. This is demonstrated in figure 3, in which there is no overlap between the confidence intervals for labials and alveolars within either the voice or aspiration contrast.

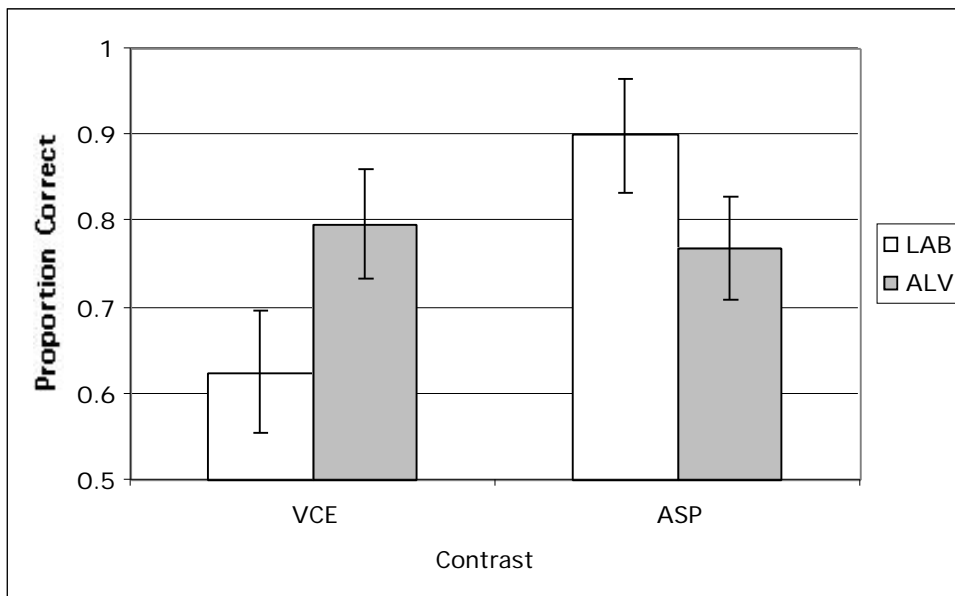


Figure 3. Results for labials vs. alveolars

In sum, in this experiment there was no evidence of a difference in discrimination of voice and aspiration across the lexical and non-lexical tasks; aspiration was discriminated better in both cases. There is, however, a difference in the discrimination of these contrasts across place of articulation: the aspiration contrast is discriminated better in labials, and the voice contrast is discriminated better in coronals. One might infer from these discrimination results that the plain (i.e. unaspirated, unvoiced) alveolar stops are perceived as relatively "voiceless", so that they are discriminated more easily from voiced stops, and with more difficulty from aspirated ones. This interpretation can be tested by examining the categorization results in the next section.



### 3. Categorization Experiment

#### 3.1. Method

The subjects in group 1 were 10 native speakers of Canadian English, undergraduate and graduate students, with no exposure to a second language beyond high school French or German, and no linguistics training. They were recruited through word of mouth, and were paid \$10 for their participation. The subjects in group 2 were 7 of the subjects in the discrimination experiment above, who participated in this experiment immediately following the previous one (three subjects were not tested due to experimenter error).

The aural stimuli were the 54 tokens used in the discrimination experiment above. Subjects in group 1 had no exposure to the stimuli before the testing phase. Subjects in group 2 were exposed to the stimuli in the training and testing sessions described in 2.1.

The alveolar- and labial-initial words were presented in separate blocks. In the alveolar block, subjects were asked to identify the initial consonant of the words as English 't' or 'd', and in the labial block they were given the choice between 'p' and 'b'. Within each block the tokens were presented in random order; the ordering of the blocks was counterbalanced across subjects. For Group 1 each of the tokens was presented 3 times; due to time constraints, Group 2 judged each token only twice.

#### 3.2 Results

Table 3 presents the results of the categorization task in terms of the proportion of voiced responses, averaged across subjects in each group, and across tokens of each voicing and place type. Mean voice onset time (VOT) for each type of stop is also indicated in milliseconds (ms). VOT was measured by independent inspection of waveform and spectrographic displays by the author and a second judge, with any differences resolved through discussion and mutual agreement. While we are reasonably confident about the accuracy of these measurements, we note that there was occasionally some ambiguity in where the onset of voicing occurred in the plain stops. Standard deviations in table 3 indicate variance across tokens.

		<b>GROUP 1</b>	<b>GROUP 2</b>	<b>VOT</b>
<b>Voiced</b>	<b>LAB</b>	0.99 (0.02)	0.98 (0.05)	-95.03 (17.27)
	<b>ALV</b>	0.99 (0.02)	0.98 (0.05)	-82.66 (18.10)
<b>Plain</b>	<b>LAB</b>	0.78 (0.21)	0.84 (0.26)	+7.39 (4.38)
	<b>ALV</b>	0.40 (0.22)	0.41 (0.25)	+11.29 (3.74)
<b>Aspirated</b>	<b>LAB</b>	0.01 (0.02)	0.02 (0.04)	+70.90 (10.83)
	<b>ALV</b>	0.03 (0.03)	0.00 (0.00)	+71.24 (13.50)

Table 3. Proportion categorized as "voiced" and VOT

Unsurprisingly, the voiced stops are almost always identified as voiced 'b' or 'd', while the aspirated stops very rarely are. In the case of the plain stops, the labials behave as might be expected from research on the perception of synthetic VOT continua. The VOT of these stops is within the range of synthetic stimuli usually labeled as 'b' by English speakers, and these natural stimuli are predominantly categorized as such. The alveolars, however, are most often identified as voiceless 't'. Paired one-tailed t-tests by subjects find the labial/alveolar difference to be highly significant for both groups (Group 1:  $t(9) = 6.52, p < .0001$ ; Group 2:  $t(6) = 3.89, p < .005$ ), as do one-tailed t-tests by items (Group 1:  $t(16) = 3.71, p < .002$ ; Group 2:  $t(16) = 3.52, p < .002$ ).

The longer VOT's of the alveolars appear to be only partially responsible for these results. A simple regression analysis with VOT and proportion of voiced responses as factors finds only a moderate negative correlation ( $r = -.348, F(1,34) = 4.70, p < .04$ ). Figure 4 plots the categorization results for both Group 1 and Group 2 against VOT.

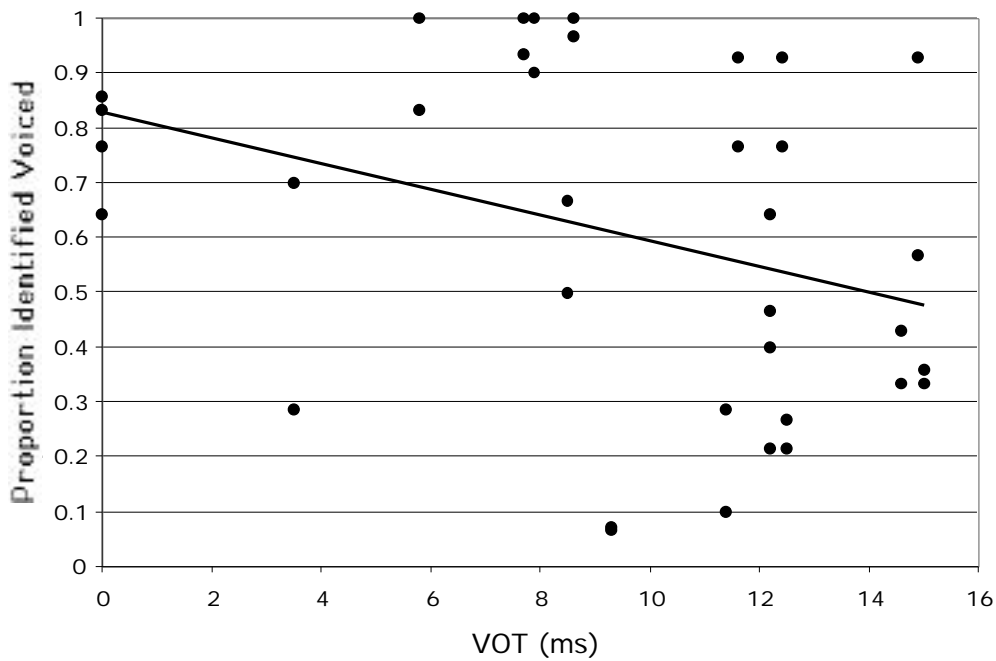


Figure 4. Categorization vs. VOT of voiceless unaspirated stops

As Figure 4 includes data points for both groups, there are twice as many points as short-lag tokens. This small number of tokens precludes detailed exploration of the relationship between acoustic cues and the listeners' judgements, but preliminary analysis of a number of candidate cues failed to reveal any strong relationships. Impressionistically, however, the plain alveolars seem to be characterized by a greater degree of "abrupt onset" of voicing (cf. Williams 1977) than the labials.

## 4. Conclusions

This study set out to replicate the lexical task effect found by Curtin *et al.*, and to further investigate the relative contributions of stimulus and task properties to the divergence of Curtin *et al.*'s results from those of earlier VOT perception research. In the attempted replication, it was found that in a lexical task (PSS), discrimination was significantly more accurate for the aspiration than the voice distinction, the opposite of Curtin *et al.*'s finding of better performance on the latter contrast in the picture selection task, even though the stimuli used in the two studies were identical. There was no difference in discrimination between tasks that required or did not require lexical access; the results of the PSS and SSS tasks here are remarkably uniform. The degraded performance on the SPP task involving minimal pairs is puzzling, but it does not speak to this issue, as discrimination of neither the voice nor the aspiration contrasts rose above chance.

The reduced ability to discriminate aspiration contrasts in the lexical task in the Curtin *et al.* study was taken to support the standard position that lexical representations are unspecified for predictable information, since aspiration is predictable in the subjects' native language. The fact that this result does not seem to be robust across experimental conditions does diminish the strength of this support. It is difficult to know exactly why the two experiments differed in the way that they did; it is perhaps significant that the Curtin *et al.* study was carried out in Montreal, where the subjects would have at least overheard the French [+/-voice] distinction (this study was undertaken in Western Canada).

One might suppose then that the present finding of good discrimination of aspiration in a lexical task could be taken as confirming the position that lexical representations are specified for predictable information, as per Lexicon Optimization. However, it is in fact questionable whether experiments of this type do speak to this question at all. Assuming that predictable information is unspecified lexically, then when English speakers are perceiving word-initial voiceless stops in their native language, they must be mapping surface aspirated stops to underlying unaspirated stops. The results of this study could be explained by making the entirely reasonable assumption that this mapping process is carried over into the second language.

The finding of better discrimination of aspiration than voicing is in line with earlier VOT research carried out with synthetic stimuli. However, in both the discrimination and categorization tasks, the plain alveolar stops seem to be consistently perceived as more voiceless than the labials; this effect was consistent across tasks and across groups of subjects. Moreover, the same place-related difference in voicing discrimination, though slightly less robust, was also reported for the subjects in the Curtin *et al.* study (Curtin 1997). The fact that this stimulus effect does not seem to be explicable in terms of VOT alone suggests that other voicing cues might play a role in English speakers' categorization of Thai stops, or that some other voicing/place interaction is involved in judgements. What those cues and interactions might be can only be determined by 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, 569-73.
- Brown, C. 1998, The role of the L1 grammar in the L2 acquisition of segmental structure.

- Second Language Research* 14, 136-93.
- Chomsky, N. and M. Halle. 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 environment for designing psychology experiments. *Behavioral Research Methods, Instruments and Computers* 25, 257-73.
- Curtin, S, Goad, H. and Pater, J. 1998: Phonological transfer and levels of representation: the perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research* 14, 389-405.
- Curtin, S. 1997: The effects of place on the perception of voicing and aspiration contrasts. Ms., University of Southern California.
- Idsardi, W. 1997: Against lexicon optimization. Talk delivered at the Montreal-Ottawa-Toronto Phonology Workshop. University of Toronto.
- Itô, J. Mester, A. and Padgett, J. 1995: Licensing and underspecification in Optimality Theory. *Linguistic Inquiry* 26: 571-613.
- Kager, R. 1999: *Optimality Theory*. Cambridge University Press.
- Lahiri, A. and Marslen-Wilson, W. 1991. The Mental Representation of Lexical Form: A Phonological Approach to the Recognition Lexicon. *Cognition*. 38, 245-94.
- Kenstowicz, M. and Kisseberth, C. 1977: *Topics in Phonological Theory*. New York: Academic Press.
- McCarthy, J. 2002: *A Thematic Guide to Optimality Theory*. Cambridge University Press.
- Ohala, J. and Ohala, M. 1995. Speech Perception and Lexical Representation: The Role of Vowel Nasalization in Hindi and English. In Connell, B. and Arvaniti, A. (editors) *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*. Cambridge University Press, 41-60.
- Prince A. and Smolensky, P. 1993: Optimality theory: constraint interaction in generative grammar. Ms., Rutgers University and University of Colorado, Boulder.
- Strange, W. 1995: Speech perception and linguistic experience: issues in cross-language research. Timonium, MD: York Press.
- Williams, L. 1977: The voicing contrast in Spanish. *Journal of Phonetics* 5, 169-184.