Appears (2008) in J.P. Avery, B.E. Dresher & K. Rice (eds) *Contrast in phonology: Theory, perception, acquisition*. Berlin: Mouton de Gruyter, pp. 323-346.

Some reflections on abstractness and the shape of inputs: The case of aspiration in English*

Heather Goad
McGill University
heather.goad@mcgill.ca

1. Preliminaries

Modern phonological theory has typically aimed to provide a unique underlying representation for a given morpheme in spite of the presence of morphophonemic alternation (cf. the historical overview in Anderson 1985). The result is a one-to-many mapping between levels of representation and, accordingly, the question of what information is present in inputs has been of central importance in theory development. While early generative phonology held the view that inputs are abstract (Chomsky and Halle 1968), the advent of Optimality Theory (Prince and Smolensky 1993/2004) has marked a shift away from this position. Although Optimality Theory includes the assumption that there are no constraints on the shapes of inputs, Lexicon Optimization guides learners in the usual case to select inputs which correspond to one of the surface forms attested in the language, that is, inputs which are not underspecified. This line of thinking has been taken a step further in the work of researchers who adopt the position that the phonetics and phonology form a single module of the grammar; inputs are phonetically enriched, inconsistent with their being underspecified (see, e.g., Boersma 1998, Steriade 2000, Flemming 2001, Curtin 2002 for proposals along these lines).

In this paper, I address the question of the shapes of inputs from the vantage point of second language acquisition. The principal goal is to determine the kind of information that is stored in native-language input representations through observing the effects of transfer from the first language into the second language. Using experimentally-obtained results on the second language acquisition of laryngeal contrasts by English learners of Thai, I will attempt to demonstrate that inputs must be abstract. Specifically, despite the presence of aspiration in the onset of stressed syllables in English, I will argue from the patterns of behaviour that emerge in the second language data that English cannot have the feature which formally marks aspiration present in inputs. A more general goal of the paper is to draw attention to the issues that the data under investigation raise concerning abstractness, in the context of current thinking in phonology.

2. Outline of the issues

Most of the empirical generalizations discussed here come from earlier collaborative work with Suzanne Curtin and Joe Pater (Curtin, Goad, and Pater 1998). Curtin, Goad, and Pater report on an experiment where English- and French-speaking subjects were taught Thai words which exploit the three-way laryngeal contrast found in this language. To provide a context for the issues to be discussed, I begin by briefly presenting the principal finding of Curtin, Goad, and Pater. When anglophones were tested using a methodology that taps lexical representations (Minimal Pair Identification task[1]), they performed significantly better on the Voiced-Plain contrast than on Plain-

Aspirated. In fact, their performance on Plain-Aspirated was poor enough to suggest that this contrast is funnelled into a single input representation, as schematized in (1) for labials.

(1) Minimal Pair task:
Stimuli: [b] [p] [p^{h}]
Identified as: /b/ /p/

In research on second language acquisition, the generally-held view is that learners initially transfer properties from their native language grammar into the second language. Accordingly, Curtin, Goad, and Pater argue that the results in (1) support the view that English speakers' inputs for Thai are underspecified for [spread glottis], the feature marking aspiration, defined as presence/(absence) of significant glottal width at the point of release of a stop. Inputs are only specified for what is contrastive in English, namely [voice], which indicates presence/(absence) of vocal cord vibration. If this is the correct interpretation of (1), it speaks against Lexicon Optimization: as voiceless stops in English are aspirated foot-initially, Lexicon Optimization will favour the input specification of [spread glottis] in this position (§4). It is also inconsistent with the view that inputs are phonetically-enriched; the latter would favour the inclusion in inputs of the set of phonetic properties which together mark aspiration. Finally, it is inconsistent with proposals which consider English to be a language in which [spread glottis] (or its equivalent) is underlyingly present and [voice] (or its equivalent) is not specified (e.g., Harris 1994, Iverson and Salmons 1995, Avery 1996).

While a logical conclusion to draw from (1) is that inputs are underspecified for [spread glottis], the validity of this interpretation is questioned when the additional results in (2) are considered; all appear to demonstrate a role for [spread glottis]:

(2)
a. In the Minimal Pair task, English speakers performed significantly better on Aspirated-Voiced than on Plain-Voiced;
b. A subset of English speakers performed well on Aspirated-Plain late in the experiment;
c. Good results on Aspirated-Plain were obtained in the ABX task, in contrast to the Minimal Pair task;
d. Good results on Aspirated-Plain were obtained in Pater's (2003) replication of Curtin, Goad, and Pater using a methodology that taps lexical representations.

My goal will be to demonstrate that the position that inputs are unspecified for [spread glottis] can be upheld, in spite of the observations in (2).

3. Curtin, Goad, and Pater's experiment

3.1. Predictions

As mentioned in §2, Thai has a three-way laryngeal contrast; both [voice] and [spread glottis] are distinctive. English and French only exhibit a two-way contrast, usually described as involving the feature [voice]. These languages differ, though, in that aspiration is absent from French but contextually present in English: voiceless stops are aspirated foot-initially ([rǽpəd]–[rəp^{h}ɪ́dəti] 'rapid'–'rapidity'). In theories of generative phonology which assume that inputs only contain contrastive material and that [voice] is the relevant distinctive feature in English, voiceless stops are underlyingly represented as unaspirated, and [spread glottis] is supplied by rule. When considering adult English speakers who are attempting to learn the three-way contrast in Thai, this approach predicts that voicing should emerge first in the interlanguage grammar; as [voice] is stored in English inputs, it should be the laryngeal feature available for transfer. Accordingly, aspirated and plain stimuli, both of which are [-voice], should initially be funnelled into a single category in contrast to voiced stimuli.

This prediction appears to be challenged by findings from the speech perception literature. As schematized in (3) for labials, when anglophones are presented with synthesized Voice Onset Time correlates of the Thai Voiced-Plain and Plain-Aspirated contrasts, they identify stimuli whose Voice Onset Time values correspond to Thai plain [p] as ‘b’, not as ‘p’ (Abramson and Lisker 1970; replicated by Strange 1972, Pisoni et al. 1982, among others).

(3) English speakers’ identification of Voice Onset Time correlates of Thai [voice] and [spread glottis]:

Stimuli: [b] [p] [p^h]

Identified as: ‘b’ ‘p’

This finding is not surprising when the Voice Onset Time values obtained by Lisker and Abramson (1964) for Thai and English are compared. Table 1 reveals that English /b, d/ align most closely with Thai /p, t/, while /p, t/ align most closely with /p^h, t^h/.[2,3]

Table 1. Voice Onset Time in msec.

Thai (3 speakers)	/b/	/p/	/p^h/	/d/	/t/	/t^h/
Average	-97	6	64	-78	9	65
Range	-165:-40	0:20	25:100	-165:-40	0:25	25:125
English (4 speakers)		/b/	/p/		/d/	/t/
Average		1	58		5	70
Range		0:5	20:120		0:25	30:105

The results in (3) demonstrate that English speakers can perceive aspiration more easily than voicing, at least in terms of Voice Onset Time. This may suggest that [spread glottis] (or the corresponding Voice Onset Time range) rather than [voice] is stored in inputs, as has recently been proposed by Harris (1994), Iverson and Salmons (1995), and Avery (1996), as mentioned above. Before underlying [voice] can be rejected, however, it is important to consider the type of methodology employed in the speech perception literature. These studies use phoneme identification and discrimination tasks which require that subjects distinguish minimally different sounds, either by labelling the sounds with orthographic symbols, or by indicating whether two sounds are the same or different. They do not require access to stored representations, as does the methodology employed by Curtin, Goad, and Pater (§3.2). Nevertheless, if the order of acquisition of stored contrasts in a second language correlates with relative perceptibility, then [spread glottis] should emerge first, contra the prediction of phonological approaches where English inputs only contain contrastive [voice].

3.2. Methodological concerns

In order to investigate the divergent predictions outlined above, Curtin, Goad, and Pater required that subjects learn 18 Thai words (6 Aspirated-Plain-Voiced minimal sets).

The main indicator of subjects' discrimination abilities was considered to be a task that taps underlying representations, the Minimal Pair task described in (4a).

(4) a. Minimal Pair task:
Subjects hear a Thai word which is the correct label for one of three pictures displayed on a computer screen. Names for two of the pictures form a minimal pair; the third is a foil. Subjects press the key which corresponds to the correct picture.

b. ABX task:
A minimal pair AB is presented aurally, followed by a third word X that matches either A or B. Subjects press the key which indicates that X is most like A or most like B.

The ABX task described in (4b), which used exactly the same stimuli as the Minimal Pair task, was also designed to tap stored representations: the tokens for A, B and X were produced by different speakers, and the interstimulus interval between B and X was relatively long. However, the methodology does not necessitate access to stored representations, a point which will be returned to in §5.3.

3.3. Minimal Pair results and interpretation

The results of Curtin, Goad, and Pater's Minimal Pair task are in Table 2. English and French speakers performed strikingly similarly on this task; indeed, in an Analysis of Variance examining contrast, language, and testing day, no effect was found for language, only for contrast.

Table 2. Proportion correct in Minimal Pair task

Testing	Aspirated-Plain		Plain-Voiced		Aspirated-Voiced	
Day	English	French	English	French	English	French
2	.59	.60	.82	.75	.93	.91
4	.63	.60	.77	.81	.95	.94
11	.68	.59	.82	.81	.95	.96

Concerning the latter, performance on Plain-Voiced was significantly better than Aspirated-Plain for both groups of learners. In fact, both groups discriminated Aspirated-Plain at only slightly better than chance.

The Minimal Pair results are not consistent with the speech perception literature which, recall, found better results on Voice Onset Time correlates of aspiration, not voice. Since correct responses on the Minimal Pair task must be made on the basis of stored representations, Curtin, Goad, and Pater maintain that the results reveal that [voice], not [spread glottis], is what English (and French) speakers transfer and thus initially represent when acquiring Thai.[4] These results support the view that learners do not first acquire the contrast that is most perceptible but, instead, that which corresponds to what many generative phonologists treat as underlying, namely [voice]. Accordingly, English inputs are underspecified for [spread glottis], despite the presence of surface aspiration in this language. The consequences of this for Lexicon Optimization are discussed next.

4. Lexicon Optimization

As mentioned in §1, Optimality Theory does not place any constraints on the shapes of inputs (what is referred to as Richness of the Base (Prince and Smolensky 1993/2004)). The burden of selecting correct outputs is placed entirely on ranking. The result is a potentially infinite set of inputs for a given output. Below, we will investigate how the learner selects appropriate input-output pairings, focussing on [spread glottis] in English.

4.1. (Under)specification of [spread glottis]

Two grammars capturing the distribution of aspiration in English are in (6).[5] The necessary constraints are first defined (informally) in (5).

(5) $_{Ft}$[SG: Voiceless stops are enhanced by aspiration foot-initially
*SG: Stops are not aspirated
IDENT-IO(SG): Correspondent segments have identical values for [spread glottis]

(6)

Grammar 1: /p^{h}æt/	$_{Ft}$[SG	*SG	IDENT (SG)
a. [pæt]	*!		*
☞ b. [p^{h}æt]		*	

Grammar 2: /pæt/	$_{Ft}$[SG	*SG	IDENT (SG)
a. [pæt]	*!		
☞ b. [p^{h}æt]		*	*

How do learners select among alternative grammars like those in (6)? Following Smith (1973), the most commonly-held view in the literature on first language acquisition is that the child's input is equivalent to the adult's output (but cf. Macken 1980, Rice and Avery 1995, Brown and Matthews 1997), until evidence to the contrary is encountered. This will lead to the child selecting Grammar 1 at Stage 1. Concerning later developmental stages, exactly what constitutes evidence to the contrary depends on the theory adopted: absence of contrast or absence of alternations. In underspecification theory, the former is (explicitly or implicitly) relevant: inputs only contain contrastive features. As aspiration does not have this status in English, it will be underlyingly unspecified, leading to selection of Grammar 2.

In Optimality Theory, by contrast, Lexicon Optimization typically steers learners toward inputs that are not underspecified: in the absence of alternations, it reconciles learners to the input-output pairing where faithfulness is maximally respected (Prince and Smolensky 1993/2004, Inkelas 1994, Itô, Mester, and Padgett 1995). In the presence of alternations, inputs may be underspecified, but only in those contexts where the alternations are observed. Since voiceless stops are aspirated foot-initially in English, Lexicon Optimization favours the specification of [spread glottis] in this position in non-alternating forms like 'pat', leading to the selection of Grammar 1. (For alternating forms like 'rapid'–'rapidity', Grammar 2 will be selected.)

The laryngeal contrasts in Curtin, Goad, and Pater's Thai stimuli were in word- and foot-initial position (and displayed no alternations). Accordingly, the presence/absence of input [spread glottis] in this position should transfer to the English learners' grammar of Thai. If [spread glottis] is specified as per Lexicon Optimization (Grammar 1), the Thai plain-aspirated contrast should be perceptible to English speakers. If [spread glottis] is underspecified (Grammar 2), English speakers should collapse plain and aspirated stimuli into a single category. Only the latter correctly predicts the asymmetry observed by Curtin, Goad, and Pater in (1): plain and aspirated stimuli are perceived as the same by anglophones, in contrast to voiced stimuli, contra the predictions of Lexicon Optimization. We attempt to resolve this problem below.

4.2. Selecting underspecified inputs

Thus far, we have discussed how the finding in (1) reveals that anglophones cannot have [spread glottis] present in inputs. Since it has already been observed that both grammars in (6), where inputs do and do not contain [spread glottis] respectively, will select the correct output in production, the challenge is for the ranking in (6) to lead to the removal of [spread glottis] in perception, appropriately resulting in underspecified inputs.

Figure 1 shows the connection between perception and production within a single grammar, as envisaged here.

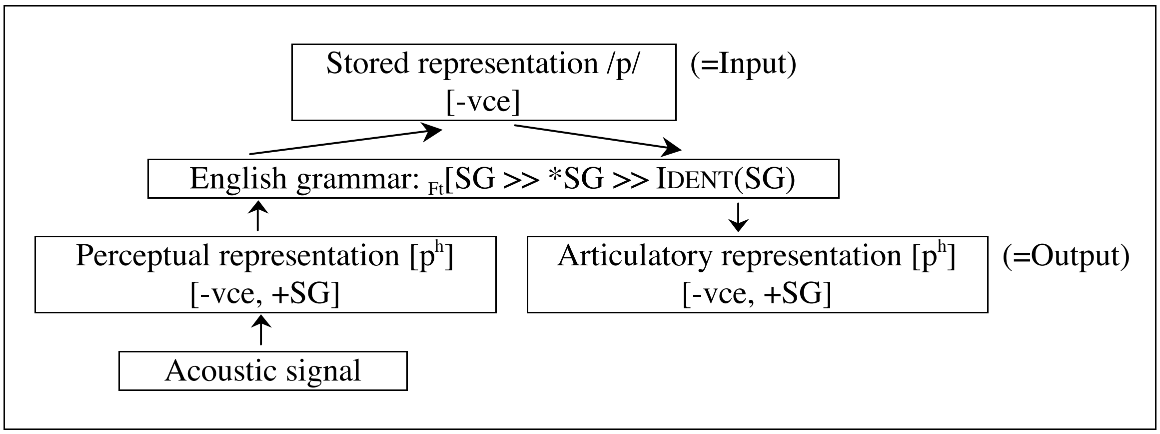

Figure 1. Perception and production in a single grammar

Focusing on perception, the processor must extract from the acoustic signal the correlates of [-voice] and [+spread glottis] which are part of the perceptual representation (Output) for [p^h]. When this form is passed up through the grammar, aspiration must be removed from [p^h], on its way to being mapped to the abstract form (Input) /p/.

I suggest that removal of aspiration occurs because of the type of constraint responsible for the presence of [spread glottis] in English. Since aspiration is contextually-determined in this language, a position-sensitive constraint, $_{Ft}$[SG, outranks *SG. Importantly, the context where [spread glottis] surfaces in English is prosodically- rather than morphologically-determined. If inputs are not prosodified, as is standardly assumed,[6] then $_{Ft}$[SG will have no impact on the shapes of inputs. Only *SG, the next constraint in the ranking, will play a role, thereby resulting in the removal of [spread glottis] from inputs, the desired result.

5. Evidence from Curtin, Goad, and Pater that aspiration is specified in inputs?

We have just seen that, by considering the type of markedness constraint involved, it is possible to select as optimal inputs which are unspecified for [spread glottis] even when outputs are uniformly aspirated. The approach was motivated by the principal finding from Curtin, Goad, and Pater from which it was concluded that English speakers (learners of Thai) cannot have [spread glottis] present underlyingly. Recall from §2, however, that there are additional results, in (2), which may lead us to question this conclusion: all of them appear to demonstrate a role for [spread glottis] in the English grammar. In the following sections, I return to these results, addressing for each whether [spread glottis] must be posited in inputs. I begin with (2a), performance on Aspirated-Voiced in the Minimal Pair task.

5.1. Aspirated-Voiced condition

Recall from §3.3 that in the Minimal Pair task, performance on Plain-Voiced was significantly better than Aspirated-Plain for both groups of learners. At that point, there was no discussion of Aspirated-Voiced; however, Table 2 reveals that performance on this contrast is near ceiling. Indeed, Aspirated-Voiced vs. Aspirated-Plain reaches a higher level of significance than Plain-Voiced vs. Aspirated-Plain. Curtin, Goad, and Pater attribute this to the observation that aspirated stops cue the voiced-voiceless contrast better than plain voiceless stops. While they specifically say that this does not indicate that both [voice] and [spread glottis] are present underlyingly, they do not address the following problem: if aspirated stops signal the voicing contrast better than plain stops, how can this information be accessible to learners if inputs, the level targeted in the Minimal Pair task, have no access to [spread glottis] (as, for example, in the model in Figure 1)?

Expressed differently, does ceiling performance on Aspirated-Voiced force [spread glottis] to be present in English inputs? The source of the answer to this lies in the performance of the francophones on the Minimal Pair task. Table 2 shows that the francophones do as well as the anglophones on Aspirated-Voiced. As [spread glottis] plays no role in the French grammar, the question cannot be reduced to the status of [spread glottis] – as allophonic – in the English grammar. Accordingly, the issue does not concern Lexicon Optimization, determining whether [spread glottis] is present in English inputs and, thus, in the transferred grammar that English speakers build for Thai. Instead, if performance on Aspirated-Voiced leads to the input specification of [spread glottis] in English, it must be present in French as well. The question thus concerns whether or not inputs are phonetically enriched. If they are, aspiration would better cue the voicing contrast because the acoustic correlates of [spread glottis] present in the signal become part of the input, independent of the language.

The numbers in Table 2 clearly reflect the fact that there is gradience in the acoustic signal, simplified somewhat, on the Voice Onset Time dimension. The gradience must map onto a set of formal objects (features), but what do these features look like? For present purposes, I will consider the two options in Figure 2. In (a), the signal is gradient, but phonological features are binary, because perception is deemed to be categorical.[7] In (b), features ([Voice Onset Time] and others) are gradient, because perception is deemed to be continuous.

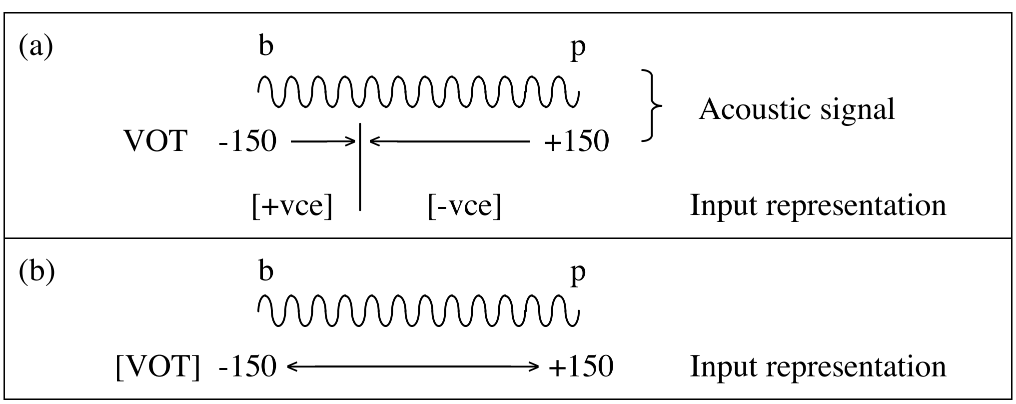

Figure 2. Input representations, using (a) binary and (b) gradient features

Given the findings from the Minimal Pair task – that Aspirated-Voiced vs. Aspirated-Plain reaches a higher level of significance than Plain-Voiced vs. Aspirated-

Plain – we might be tempted to conclude that perception is continuous and must be reflected in the grammar as in (b) in Figure 2. To assess this, we turn briefly to consider the research on Categorical Perception. Repp (1984: 251-252) defines Categorical Perception as “the experience of discontinuity as a continuously changing series of stimuli crosses a category boundary, together with the absence of clearly perceived changes within a category”. In the perception of speech, this research has looked at the perceptual reality of discrete segments which (more or less) correspond to phonemes.

Concerning voicing in stops, Categorical Perception effects are particularly robust. While one might thus be tempted to conclude that (a) in Figure 2 is correct, there is also a large literature which has found that perception can be continuous (see Repp 1984 for a review). This work has focussed on determining the experimental conditions that can be manipulated to lead to either categorical or continuous perception. Does this research argue against perception as categorical and thus in favour of (b) in Figure 2? The answer, I believe, is no. What it does show is that while Categorical Perception effects are widely observed, the strongest version of the Categorical Perception hypothesis cannot be maintained, as there are experimental conditions under which listeners can discriminate within-category differences.

At this point, one might conclude that a decision between (a) and (b) in Figure 2 cannot be made. It is not obvious, however, how (b) would predict Categorical Perception effects at all, whereas (a) does allow for diversions from Categorical Perception. To explore how (a) permits such diversions, we turn to consider the different processing levels proposed by Werker and Logan (1985). Werker and Logan demonstrate that listeners can exploit different processing strategies, depending on experimental conditions, especially interstimulus interval, and also practice gained during the experiment itself. See (7):

(7)
a. Phonemic: Stimuli perceived according to native language phonemic categories;
b. Phonetic: Sub-phonemic information perceived;
c. Acoustic: Finer acoustic detail between stimuli perceived.

Let us consider (7a-b) in the context of Figure 1 above. Phonemic processing will only access what is available in the stored representation; phonetic processing will access non-contrastive information as well, available in the perceptual representation. The essential point, then, is that while experiments can be designed to tap different levels of representation, stimuli are funnelled into native phonemic categories, once the information available in the phonetic code has decayed. Accordingly, there must be a level of representation that reflects the type of information that is perceived under such conditions – the Input in Figure 1.

Since Curtin, Goad, and Pater’s Minimal Pair task requires access to inputs, it must involve phonemic processing. The results should therefore support the Categorical Perception hypothesis, (a) in Figure 2. I believe that they do. Recall from Table 2 that Aspirated-Plain was discriminated only slightly better than chance. This indicates that these stimuli form one category, [-voice]; however, some members of this category, the aspirates, are better instances of [-voice] than other members, resulting in ceiling performance on Aspirated-Voiced. In short, while some types of information in the acoustic signal (the phonetic correlates of aspiration) cue the voiced-voiceless contrast particularly well, poor performance on Aspirated-Plain strongly suggests that this information is not encoded in inputs.

5.2. Performance on day 11

In this section, we turn to examine the performance on Aspirated-Plain at Day 11 where some improvement is observed among the anglophones (see (2b)).[8] The overarching question, as before, is whether these results demonstrate a role for [spread glottis] in inputs.

One question posed by Curtin, Goad, and Pater is whether surface aspiration in English has any positive effect on speakers' ability to underlyingly represent this feature. Recall from §3.3 that an Analysis of Variance did not find the improvement on aspiration observed at Day 11 to be significant. To further explore the issue of whether the improvement reflected genuine development, Curtin, Goad, and Pater subtracted the participants' Day 2 scores from their Day 11 scores and subjected the scores to a Kolmogorov-Smirnov test. The difference between the anglophones and francophones was significant. However, development was only observed for three anglophones: they showed an average improvement of 24%; the remaining five showed no improvement overall. A Kolmogorov-Smirnov test considered these two groups of anglophones to be significantly different.

Do these results suggest that [spread glottis] is present in native English inputs? The answer, I argue, is no. First, the presence of [spread glottis] – as mandated by Lexicon Optimization – cannot account for the observation that on Days 2 and 4, the anglophones only performed slightly above chance on Aspirated-Plain. Second, their performance on Days 2 and 4 is the same as the francophones who do not have [spread glottis] in their grammar. Finally, as just mentioned, the improvement at Day 11 is only observed for a subset of anglophones.

The presence of surface aspiration in English can have an effect on speakers' ability to *eventually* store this feature in their second language inputs. Indeed, the findings for Day 11 suggest that [spread glottis] has truly been phonologized in the grammars of the anglophone individuals involved. However, the presence of surface aspiration cannot, I suggest, have an effect at the outset of acquisition. The developmental scenario for second language acquisition is outlined in Figure 3. Stage 1 (Days 2 and 4) represents the transferred English grammar. The feature [spread glottis] has the same status as in the native English grammar: it is absent from inputs. Because of the two-way contrast in voicing in the transferred grammar, Thai [p] and [p^h] are mapped to a single category /p/. It is hypothesized that outputs will show aspiration for target [p] and [p^h], due to high-ranking $_{Ft}$[SG (production was not tested). Stage 2 reflects the development exhibited by the three anglophones (Day 11). The three-way contrast is now perceptible as reflected by the demotion of *SG below the faithfulness constraint IDENT(SG). Without demotion of $_{Ft}$[SG, production outputs are, for all intents and purposes, unaffected. This developmental path, that production lags behind perception, is commonly observed in first language development.

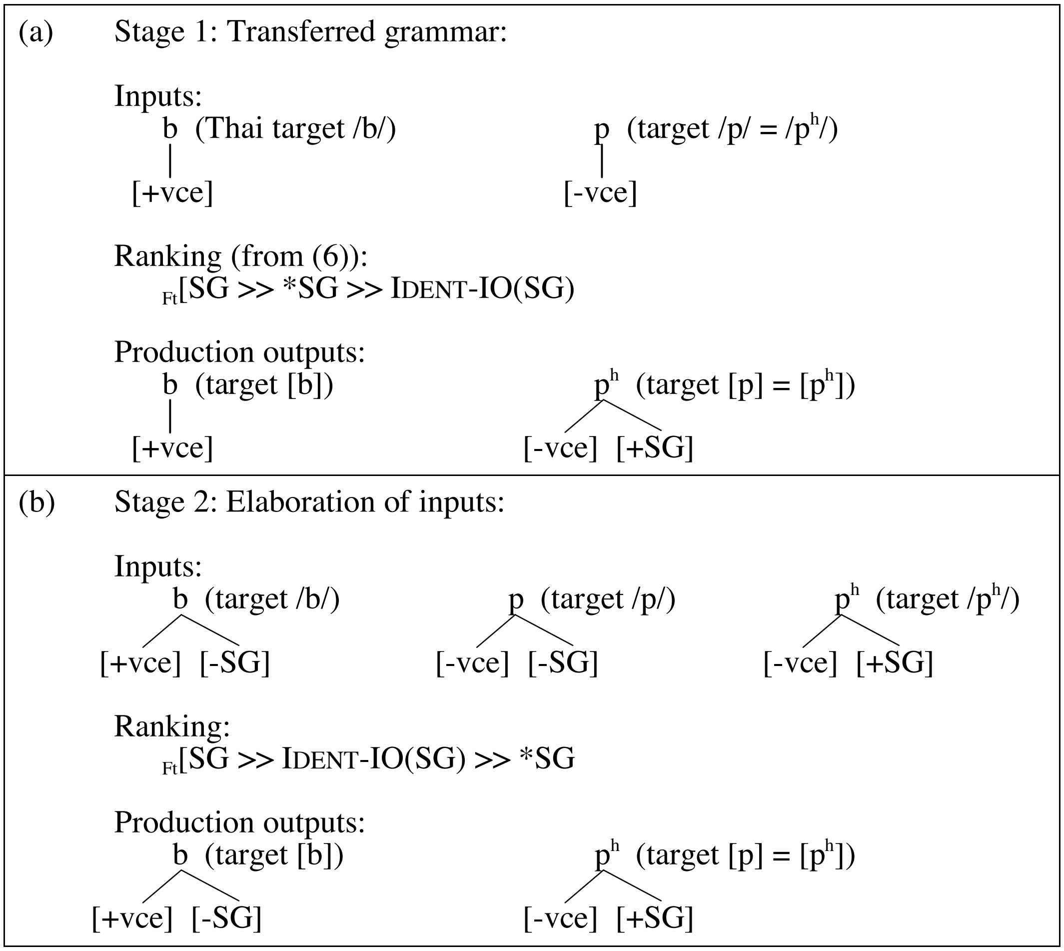

Figure 3. Stages in development

Three principal claims are being made here. One, development over time in Optimality Theory involves the elaboration of inputs (Goad and Rose 2004), not just constraint reranking. Two, the lexicalization of new features can only occur over time. Indeed, there were no English speakers in the Curtin, Goad, and Pater study who were able to perceive the Thai three-way voicing contrast from the outset. Three, there is a relationship between the presence of allophonic aspiration in the native language and the ability to lexicalize this feature in the second language. This is in the spirit of Brown (1998) but represents a weakening of her proposal. Brown hypothesizes that beyond the transfer stage, only features which are contrastive in the native language grammar can be combined to build new segments in a second language. This proposal is being extended here to include non-contrastive features.

As [spread glottis] has no status in French, the predictions made for this population of speakers are the same as Brown: the Thai three-way contrast should never be lexicalized. That is, in Figure 1, [spread glottis] will never be mapped from the acoustic signal into the perceptual representation. Whether or not this prediction can be upheld remains to be investigated.

5.3. Curtin, Goad, and Pater's ABX results

In this section, I address the third issue concerning the role of [spread glottis] in the grammar transferred from English to Thai, that better results on [spread glottis] were obtained on Curtin, Goad, and Pater's ABX task than on their Minimal Pair task (see (2c)). Compare Table 3 with Table 2 from §3.3.

Table 3. Proportion correct in ABX task

Testing	Aspirated-Plain		Plain-Voiced		Aspirated-Voiced	
Day	English	French	English	French	English	French
2	.84	.64	.83	.78	.99	.96
4	.77	.67	.73	.77	.99	.98
11	.79	.59	.70	.88	.88	.98

What is striking about these results, when compared with the Minimal Pair results, is that there are differences across languages in the Aspirated-Plain and Plain-Voiced conditions: francophones performed better on Plain-Voiced than on Aspirated-Plain, as they did in the Minimal Pair task, but anglophones performed similarly on these two contrasts, unlike in the Minimal Pair task. Aspirated-Plain vs. Plain-Voiced was significant for the francophones only; thus, while the numbers in Table 3 may suggest that the anglophones are performing better on Aspirated-Plain than on Plain-Voiced, this is not significant.

Why do the anglophones perform better on Plain-Voiced than on Aspirated-Plain in the Minimal Pair task, but not in the ABX? And while the ABX was designed to tap inputs, performance on Aspirated-Plain is much better than expected if [spread glottis] is not underlyingly specified; does this finding suggest that [spread glottis] is present in inputs?

Although the ABX task was designed to tap inputs, the methodology does not *require* lexical access, as subjects are presented with auditory stimuli only; thus, judgements can be based on phonetic similarity alone. In Curtin, Goad, and Pater, we suggested that the results on this task were due to subjects sometimes relying on their lexical representations ([±voice]) and sometimes on surface representations ([±spread glottis]). Given the position-sensitive nature of voicing and aspiration in English, we did not consider the possibility that tapping surface representations could result in a *three*-way distinction. That is, we did not consider the possibility that speakers might process stimuli in the ABX at Werker and Logan's (1985) phonetic level (7b), where within-category decisions can be made. The means in Table 4 suggest perception of a three-way contrast: performance on *both* Aspirated-Plain and Plain-Voiced in the ABX is as good as performance on Plain-Voiced in the Minimal Pair task.

Table 4. Anglophone means in ABX and Minimal Pair tasks

	ABX		Minimal Pair	
Aspirated-Plain	.80	not significant	.60	significant
Plain-Voiced	.75		.80	

Two questions arise at this point: (i) Are Curtin, Goad, and Pater correct in concluding that the ABX is sometimes tapping lexical representations and sometimes surface representations? (ii) Do the ABX results suggest that [spread glottis] is present in English inputs? I believe that the answer to both questions is no. Concerning (i), the ABX methodology is not well-suited to eliciting phonemic judgements; it favours within-category processing, even when the experiment is designed to elicit cross-category judgements (Werker and Logan 1985, Brannen 2002). In short, the ABX methodology enables listeners to perceive the three-way Aspirated-Plain-Voiced distinction. Following from this, concerning question (ii), the results do not indicate that [spread glottis] is present in English inputs: as we have just suggested, this task is not tapping inputs.

6. Evidence from Pater's replication that aspiration is specified in inputs?

Thus far, three potential sources of evidence for the input specification of [spread glottis] in English have been examined from the results obtained by Curtin, Goad, and Pater. It has been argued for each that, counter to appearance, [spread glottis] is not present in inputs. In this section, we turn finally to Pater's (2003) replication of Curtin, Goad, and Pater which found better results for Aspirated-Plain than Plain-Voiced on a task that taps lexical representations (see (2d)). These results appear to require that [spread glottis] be specified in English inputs, contra the conclusion reached so far.

In Curtin, Goad, and Pater's study, the Minimal Pair and ABX tasks were methodologically quite different from each other. Pater attempted to rectify this by modifying the methodology as in (8). (All subjects were anglophones; stimuli were the same as in Curtin, Goad, and Pater.)

(8) XAB discrimination tasks (Pater 2003):
- a. Sound-Sound-Sound
- b. Picture-Sound-Sound
- c. Sound-Picture-Picture

Sound-Sound-Sound is most like Curtin, Goad, and Pater's ABX task, while Sound-Picture-Picture is most like their Minimal Pair task. Picture-Sound-Sound and Sound-Picture-Picture both require lexical access.

The results, averaged across subjects, are in Table 5.

Table 5. Means in Pater's XAB tasks

	Sound-Sound-Sound	Picture-Sound-Sound	Sound-Picture-Picture
Aspirated-Plain	.84	.83	.52
Plain-Voiced	.71	.72	.53

The most conspicuous result is that subjects performed only at chance on Sound-Picture-Picture. Pater is puzzled by this and thus excludes the task from further discussion; I return to this below. Second, performance is the same on both Sound-Sound-Sound and Picture-Sound-Sound, even though only the latter requires lexical access. Finally, Aspirated-Plain is significantly better than Plain-Voiced on both Sound-Sound-Sound and Picture-Sound-Sound.

A comparison of Tables 4 and 5 reveals two striking differences between Pater's and Curtin, Goad, and Pater's results. First, Pater's Picture-Sound-Sound most closely parallels Curtin, Goad, and Pater's ABX results; better performance is observed on Aspirated-Plain. As Picture-Sound-Sound requires access to inputs, we must consider whether Pater's results indicate that [spread glottis] is stored. Second, neither of Pater's tasks which require lexical access, Picture-Sound-Sound and Sound-Picture-Picture, mirror the results of Curtin, Goad, and Pater's Minimal Pair task – better performance on Plain-Voiced than on Aspirated-Plain which Curtin, Goad, and Pater use to argue against input [spread glottis].

In the following lines, I suggest that these differences arise from methodological considerations, that Pater's study is not a true replication of Curtin, Goad, and Pater. I hypothesize further that the Sound-Picture-Picture results indicate that [spread glottis] is not stored in inputs, at least not in the compositional way that native speakers store features (see below).

I begin with the duration of the experiment. Pater mentions that subjects were trained one day and tested the next. In Curtin, Goad, and Pater, subjects were similarly tested for the first time on Day 2. However, Curtin, Goad, and Pater also included a pre-test (Day 0) where subjects were tested on 18 different Thai stimuli. Although subjects were not taught the meanings of these words, they were given positive feedback on discrimination tasks. This additional exposure to Thai may have helped learners establish native-like representations for these segments.

In this context, one must question whether the subjects in Pater's experiment had enough opportunity to truly learn the words – to store them using the same set of primitives available to end-state grammars. In Sound-Picture-Picture, where performance was at chance, Pater mentions that on the foils, subjects performed near ceiling; accordingly, he concludes that they did learn the words. However, there are several cues to distinguish foils from test items; the former differed from the latter in the initial consonant's place of articulation and for at least one segment in the rhyme (all stimuli were Consonant-Vowel-Consonant in shape).

While excellent performance on the foils reveals that they are stored differently from the test stimuli, it does not tell us *how* the various stimuli are stored. We turn to this issue now. In the acquisition literature, a distinction is commonly drawn between holistic and analytic learning (e.g., Cruttenden 1981, Peters 1983 on first language acquisition; Wray 2002 on second language acquisition). An important theme that emerges from this literature is that holistic learning is common at the earliest stages in acquisition. This observation is extended to the present context as follows: at the immediate onset of perception in a second language, transfer is not yet a consideration, as stimuli are stored in holistic rather than analytic form. That is, as much information as can be extracted from the acoustic signal is stored, but this information is not yet mapped to a set of formal objects of analysis (features).

Two results suggest that a holistic, rather than compositional, analysis has been undertaken by the subjects in Pater's experiment. First, recall that performance on Sound-Picture-Picture is at chance. If subjects have not yet undertaken a featural analysis of the stimuli, good performance will require a comparison of at least two auditory stimuli, rather than an assessment based on a single stimulus as in Sound-Picture-Picture. Contrastingly, if subjects have had time to analyse and store the stimuli featurally, such a comparison will not be necessary: in Curtin, Goad, and Pater's Minimal Pair task, subjects were presented with a single auditory stimulus, and performance on one contrast, Plain-Voiced, was significantly better than chance. It must still be explained why, on the holistic view, Aspirated-Plain is perceived more accurately than Plain-Voiced on Pater's Sound-Sound-Sound and Picture-Sound-Sound tasks. This, I believe, follows from the observation that, ceteris paribus, [spread glottis] is more perceptible than [voice] (§3.1).[9]

The second result which suggests that the subjects in Pater's experiment have undertaken a holistic analysis is that there is a strong effect for place. Table 6 shows that, for both Sound-Sound-Sound and Picture-Sound-Sound, discrimination of aspiration is better for labials than for alveolars, while discrimination of voice is better for alveolars than for labials (both are significant).

Table 6. Means in Pater's tasks by place

	Sound-Sound-Sound		Picture-Sound-Sound	
	Labial	Alveolar	Labial	Alveolar
Plain-Voiced	.63	.78	.62	.81
Aspirated-Plain	.90	.78	.90	.77

If speakers have done an abstract featural analysis and display phonemic processing, place effects should not be found. Since phonemic processing accesses representations at the level of contrast, these representations will contain features for place, features for voicing, and their combinatorial possibilities, but differences in degree of voicing which are sensitive to place of articulation will not be accessible.[10] Place effects should only be present under phonetic processing which accesses non-contrastive information available in the perceptual representation, or under a holistic analysis where small differences in degree of voicing observed for different places of articulation will be stored.

If this approach is correct, place effects should be present in Curtin, Goad, and Pater's ABX task but not in their Minimal Pair task. To investigate this, we turn to Curtin (1997). Curtin observed place effects in the data collected by Curtin, Goad, and Pater, but they were largely dependent on task.

Table 7. Means in Curtin, Goad, and Pater's tasks by place

	ABX				Minimal Pair			
	English		French		English		French	
	Labial	Alveolar	Labial	Alveolar	Labial	Alveolar	Labial	Alveolar
Plain-Voiced	.64	.86	.69	.92	.75	.87	.75	.84
Aspirated-Plain	.93	.67	.73	.54	.65	.63	.60	.58

Table 7 reveals that, as expected, in the ABX, place effects were robust: performance on Aspirated-Plain was significantly better for labials than for alveolars for both language groups; Plain-Voiced exhibited the opposite pattern, with significantly better performance for alveolars. As expected, in the Minimal Pair task, labial does not enhance the perception of Aspirated-Plain, in contrast to the ABX. Unexpectedly, though, there were place effects for alveolars, with both groups performing significantly better on alveolar in the Plain-Voiced condition. Importantly, however, in contrast to the ABX, no particular place enhances the perception of Aspirated-Plain; this is consistent with the proposal that [spread glottis] is not present underlyingly. In short, the results for place are in the right direction: place effects are stronger in the ABX than in the Minimal Pair task; and perception of Aspirated-Plain is not enhanced by place in the latter.

7. Conclusion

In this paper, I have argued that inputs are abstract and, thus, that the phonology (i.e., stored representations) does not necessarily align with the phonetics. Following from this, once there has been sufficient exposure to a second language, learners' inputs will show effects of transfer where their inputs are shaped by what is stored in the first language grammar. In the present case, inputs for English learners of Thai are specified for [voice] only, not [spread glottis], as revealed by Curtin, Goad, and Pater's Minimal Pair task.

Three sources of evidence which challenge the view that [spread glottis] is absent from English inputs were examined from Curtin, Goad, and Pater's results; for each, it was argued that, counter to appearance, [spread glottis] is not underlingly specified. First, although in the Minimal Pair task, Aspirated-Voiced was perceived better than Plain-Voiced, it was argued that this reflects gradience in the acoustic signal, where this gradience maps onto abstractly-represented features, leading to categorical perception effects.

Second, concerning the acquisition of non-contrastive features like [spread glottis], it was argued that the presence of such features in native language outputs can aid in their eventual lexicalization in a second language. However, the lexicalization of such features can only be observed at non-initial stages in acquisition, consistent with [spread glottis] being absent from transferred English inputs.

Third, good performance on pairs of stimuli involving features which are not contrastive can be observed under certain experimental conditions, but this does not lead to the conclusion that such features must be stored. Specifically, better performance for anglophones on Aspirated-Plain on Curtin, Goad, and Pater's ABX task than on the Minimal Pair task does not indicate that [spread glottis] is stored. The ABX task involves phonetic processing where within-category effects are expected, leading to across-the-board good performance.

Similar across-the-board good performance was argued to have been observed for learners who have stored stimuli as featurally-unanalysed, in tasks that involve a comparison between at least two auditory stimuli, as in Pater's Picture-Sound-Sound and Sound-Sound-Sound. For such learners, it was argued to follow that poor performance will be observed on tasks where subjects are exposed to one auditory stimulus only, as in Pater's Sound-Picture-Picture. Additional stimulus effects, such as an interaction between voicing and place, were argued to be expected when stimuli are stored in holistic fashion.

A remaining question that has been left largely unaddressed concerns the weighting of the phonetic cues to onset voicing present in the Thai stimuli. If the hypothesis advanced in this paper proves to be correct, that representations are abstract, it is still of course the case that some cue or cues must have led to the profile of results obtained, notably that the anglophones in Curtin, Goad, and Pater's Minimal Pair task group together plain and aspirated stops in contrast to voiced stops. Although burst intensity leads to the right results for this task (see note 4), this is not the case for Curtin, Goad, and Pater's ABX task nor for any of Pater's tasks, where markedly different results were found. How do methodological considerations interact with the particulars of the stimuli employed to lead to the various different patterns of behaviour obtained? I leave this question to future research.

References

Abramson, Arthur and Leigh Lisker (1970) Discriminability along the voicing continuum. In: Bohuslav Hala, Milan Romportl and Premysl Janota (eds.), *Proceedings of the Sixth International Congress of Phonetic Sciences,* 569-573. Prague: Academia.

Anderson, Stephen (1985) *Phonology in the Twentieth Century*. Chicago: University of Chicago Press.

Avery, J. Peter (1996) The representation of voicing contrasts. Ph.D. dissertation, Department of Linguistics, University of Toronto.

Boersma, Paul (1998) *Functional Phonology*. The Hague: Holland Academic Graphics.

Brannen, Kathleen (2002) The role of perception in differential substitution. *Canadian Journal of Linguistics* 47: 1-46.

Brown, Cynthia (1998) The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research* 14: 136-193.

Brown, Cynthia and John Matthews (1997) The role of feature geometry in the development of phonemic contrasts. In: S.J. Hannahs and Martha Young-Scholten (eds.), *Focus on Phonological Acquisition,* 67-112. Amsterdam: John Benjamins.

Chomsky, Noam and Morris Halle (1968) *The Sound Pattern of English*. New York: Harper & Row.

Cruttenden, Alan (1981) Item-learning and system-learning. *Journal of Psycholinguistic Research* 10: 79-88.

Curtin, Suzanne (1997) The effects of place on the perception of voicing and aspiration contrasts. Ms., Department of Linguistics, University of Southern California.

Curtin, Suzanne (2002) Representational richness in phonological development. Ph.D. dissertation, Department of Linguistics, University of Southern California.

Curtin, Suzanne, Heather Goad and Joe Pater (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.

Flemming, Edward (2001) Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18: 7-44.

Goad, Heather (2000) The acquisition of voicing and aspiration contrasts by English and French learners of Thai: What's wrong with VOT? Paper presented at Department of Phonetics and Linguistics, University College London, March.

Goad, Heather and Yvan Rose (2004) Input elaboration, head faithfulness and evidence for representation in the acquisition of left-edge clusters in West Germanic. In: René Kager, Joe Pater and Wim Zonneveld (eds.), *Constraints in Phonological Acquisition,* 109-157. Cambridge: Cambridge University Press.

Harris, John (1994) *English Sound Structure*. Oxford: Blackwell.

Inkelas, Sharon (1994) The consequences of optimization for underspecification. Ms., Department of Linguistics, University of California, Berkeley.

Itô, Junko, R. Armin Mester and Jaye Padgett (1995) Licensing and underspecification in Optimality Theory. *Linguistic Inquiry* 26: 571-613.

Iverson, Gregory and Joseph Salmons (1995) Aspiration and laryngeal representation in Germanic. *Phonology* 12: 369-396.

Lisker, Leigh and Arthur Abramson (1964) A cross-language study of voicing in initial stops. *Word* 20: 384-422.

Macken, Marlys (1980) The child's lexical representation: The 'puzzle-puddle-pickle' evidence. *Journal of Linguistics* 16: 1-17.

Pater, Joe (2003) The perceptual acquisition of Thai phonology by English speakers: Task and stimulus effects. *Second Language Research* 19: 209-223.

Peters, Ann (1983) *The Units of Language Acquisition*. Cambridge: Cambridge University Press.

Pisoni, David, Richard Aslin, Alan Perey and Beth Hennessy (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, Alan and Paul Smolensky (1993) *Optimality Theory: Constraint interaction in generative grammar*. Ms., Rutgers University and University of Colorado. Published Oxford: Blackwell [2004].

Repp, Bruno H. (1984) Categorical perception: Issues, methods, findings. In: Norman J. Lass (ed.), *Speech and Language: Advances in Basic Research and Practice,* Volume 10, 243-335. New York: Academic.

Rice, Keren and J. Peter Avery (1995) Variability in a deterministic model of language acquisition. In: John Archibald (ed.), *Phonological Acquisition and Phonological Theory,* 23-42. Hillsdale, NJ: Erlbaum.

Smith, Neil (1973) *The Acquisition of Phonology*. Cambridge: Cambridge University Press.

Steriade, Donca (2000) Paradigm uniformity and the phonetics-phonology boundary. In: Michael Broe and Janet Pierrehumbert (eds.), *Papers in Laboratory Phonology V,* 313-334. Cambridge: Cambridge University Press.

Strange, Winifred (1972) The effects of training on the perception of synthetic speech sounds: Voice Onset Time. Ph.D. dissertation, Department of Speech-Language-Hearing Sciences, University of Minnesota.

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

Wray, Alison (2002) *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

* I would like to thank members of the audience at the Second International Conference on Contrast in Phonology for questions and comments on an earlier version of this paper. The paper has also benefitted from discussions with Jonathan Bobaljik, Kathleen Brannen, Suzanne Curtin, Elan Dresher, Jill Heather Flegg, Joe Pater, and Linda Polka and from comments from two anonymous reviewers. I take full responsibility for the content; in particular, Curtin and Pater are in no way responsible for the (sometimes outlandish) interpretations of the data from Curtin, Goad, and Pater (1998) and Pater (2003). This research was supported by grants from SSHRC and FQRSC.

1 In this task, subjects hear a word which is the correct name for one of three pictures. Names for two of the pictures form a minimal pair while the third is a foil. Subjects must select the picture which corresponds to the auditory stimulus (see §3.2 for further details).

2 Note that Thai has no /g/; thus, the focus of the discussion throughout this paper is on labial and coronal stops only.

3 The English values for /b, d/ in Table 1 come from 114 tokens produced almost exclusively by three of the four speakers in Lisker and Abramson. The fourth speaker produced virtually all of his voiced stops with voicing lead and was responsible for 95% of all cases of voicing lead in the sample.

4 Importantly, the Voice Onset Time values of the Thai stops in Curtin, Goad, and Pater are comparable to those of Lisker and Abramson (1964). The question that arises is what phonetic cue(s), other than Voice Onset Time, is particularly prominent in the Thai stimuli which leads speakers to group together plain and aspirated stops in contrast to voiced stops. When the Thai stimuli were examined for burst intensity (Goad 2000), it was found that the voiced stops have much bigger bursts than the plain and aspirated stops and so this is a likely candidate (average for labial and coronal voiced stops: .103 (RMS, expressed in Pascals); plain stops: .043; aspirated stops: .043). While speakers' sensitivity to burst intensity can account for Curtin, Goad, and Pater's Minimal Pair task results, it cannot account for their ABX results where good performance was observed on *both* Aspirated-Plain and Plain-Voiced (§5.3). This suggests that methodological considerations, rather than particulars of the stimuli employed, are responsible for Curtin, Goad, and Pater's Minimal Pair results. This question, however, clearly requires further examination.

5 The ranking *SG >> IDENT-IO(SG) is not evident from (6). It emerges when voiceless stops surface as plain. For example, to ensure that an input like /hæp^{h}i/ (permitted by Richness of the Base) surfaces as [hǽpi], *SG must be dominant.

6 While this is the standard position, it is counter to what is argued for in Goad and Rose (2004); at this point, I do not know how to resolve this.

7 In (a), the cross-over point between 'b' and 'p' is given as -30, following Abramson and Lisker's (1970) results for Thai speakers. This is somewhat misleading, as their results were arrived at through phoneme discrimination and identification tasks (§3.1). As we are discussing underlying representations for English speakers, the appropriate boundary should be determined using tasks that tap inputs.

8 Day 11 is one week after no exposure to Thai.

9 As an anonymous reviewer points out, it could also be that the plain stops are being perceived as voiced in these tasks. Indeed, this could perhaps lead to an explanation of why performance on Aspirated-Plain versus Plain-Voiced in Pater's Sound-Sound-Sound and Picture-Sound-Sound tasks was significant, while performance on these same contrasts in Curtin, Goad, and Pater's ABX task was not.

10 As an anonymous reviewer points out, this position may be too strong; for example, aspiration is much more salient in velars than in labials. If this perceptual effect has *phonological* consequences, then my position will have to be weakened. One possible phonological consequence would involve a language where /k/ has been singled out for spirantization, if this type of process arises from one noise source (burst) being misperceived as another (turbulence).