1150

Original Paper

Phonetica 1999;56:111–122

Received: December 18, 1998
Accepted: August 11, 1999

Thai Final Stops: Cross-Language Perception

Arthur S. Abramson[a] Kalaya Tingsabadh[b]

[a]Haskins Laboratories, New Haven, Conn., USA; [b]Department of Linguistics, Chulalongkorn University, Bangkok, Thailand

Abstract

As part of a project seeking a better understanding of the links between speech production and perception, we have conducted experiments on the perception of Thai word-final stop consonants by native speakers of Thai and native speakers of American English. The final stops of Thai are never released audibly. In English, however, final stops occur with and without audible releases. Previous work has shown that released stops are likely to be more intelligible than unreleased ones. That is, identification is better when information as to place of articulation in the formant transitions of the closing articulatory gesture is supplemented by acoustic information in a release burst. Both languages have labial, alveolar, and velar places of articulation, but Thai also has glottal closure. Two native speakers of Thai recorded sets of monosyllabic words minimally distinguished by the four final stops /p t k ʔ/. Randomized lists of the isolated words were presented to 19 native speakers of English who responded with the labels 'p', 't', 'k', and 'other'. They were also presented to 30 native speakers of Thai for identification in Thai script as words. The highly significant differences between the two groups imply the possibility that the closing gestures of the Thai stops include a component that compensates somewhat for the absence of release, a component to which the English speakers are perhaps less well attuned. Such an interference will be tested in future work by tracking of the gestures in both languages with the Haskins Laboratories magnetometer.

Introduction

This study is part of a body of research being carried out by a group of investigators at the Haskins Laboratories seeking to learn more about links between the production and perception of speech. We are trying, acoustically and physiologically, to address the validity of recent theoretical arguments that put forward the articulatory gesture, rather than components of the acoustic signal itself, as the basic unit of speech perception [Fowler, 1986, 1994; Liberman and Mattingly, 1985, 1989]. That is, for us the obvious fact that all human speech signals are generated by activities within the vocal tract is not enough to support such arguments, since that alone does not justify rejection of the claim that the objects of speech perception are purely auditory/acoustic, with articulation merely serving to provide suitable acoustic signals for the identification of phonetic entities [Diehl and Kluender, 1989].

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/99/0564–0111
$17.50/0
Accessible online at:
www.karger.com/journals/pho

Arthur S. Abramson
Department of Linguistics, U-145
University of Connecticut, Storrs, CT 06269-1145 (USA)
Tel. +1 (860) 429-1141, Fax +1 (860) 486-0197
E.Mail abramson@uconnvm.uconn.edu

One possible approach to the undertaking is to examine the case of two languages that are rather different phonologically yet have a few phonemes that, in a given context, are manifested phonetically in a very similar or even a seemingly identical manner. If, despite the apparent great similarity, there is nevertheless a critical gestural difference between the languages in the formation of these sounds, the theory would predict perceptual consequences.

Thai and English both have word-final voiceless stop consonants at three oral places of articulation that can be phonetically very similar. For example, a bilingual speaker standing on a street in Bangkok who happens to hear a passerby say, with a slightly falling intonation, the monosyllable [hɑt̚] with no audible release of the final [t] might not know, without being aware of the sociolinguistic or pragmatic context of the utterance, or phonotactic constraints imposed by its occurrence in isolation, whether the person is saying the Thai word /hàt/ 'to practice' or American English *hot*. There is, however, an important difference between the languages in the possible realizations of these stops. In English the final stops may be audibly released or not. This has no phonologically distinctive value; releases may occur randomly or optionally, the latter, perhaps, for more deliberate or emphatic speech. Crystal and House [1988] present extensive data on this from the spectrographic examination of passages of running text read aloud. For example, before silence they found release bursts for about half the sampling of word-final stops. For voiceless stops it was 55% of 111 tokens, and for voiced stops 33% of 52 tokens [Crystal and House, 1988, table 6]. In Thai, however, the final stops are not released audibly. We must stress *audibly* here. Obviously, a stop closure has to be released sooner or later. If the closure is utterance-final or, perhaps, prepausal, the supraglottal vocal tract will move into a neutral rest position. If utterance-medial, the articulators will have to move into new configurations for the continuation of speech. Although we know of no experimental study of the topic, Thai is known for the inaudibility of such releases [Noss, 1964; Abramson, 1972; Harris, 1987; Tingsabadh and Abramson, 1993]. Indeed, one of the hallmarks of a foreign accent to the Thai ear is the occasional or, worse yet, frequent occurrence of audible releases of Thai word-final stops.

Thai also differs from English in that English has both voiced and voiceless final stops, while in Thai the neutralization of the three-way contrast of voicing and aspiration in that position [Abramson, 1962, p. 4] yields only voiceless stops [Abramson, 1972; Tingsabadh and Abramson, 1993]. It has been observed [Harris, 1987, p. 55; in press] that this neutralization is implemented by glottal closure simultaneous with the oral closure.

Thai has in common with English the labial, alveolar, and velar places of articulation. A more descriptive term for the Thai alveolar place of articulation might be dentialveolar [Harris, in press]. It differs from English in also having a fourth place, glottal, as well as distinctively short and long vowels. The final glottal stop occurs lexically only after short vowels. Thus we have in words with short vowels such minimal sets as /láp/ 'to sharpen', /lát/ 'to take a shortcut', /lák/ 'to steal', and /láʔ/ 'to abandon'. We expected this striking difference between the stop systems of the languages to be useful in our experiments.

Information on the place of articulation of a stop is available in both the closing and opening gestures of the articulator; it is revealed acoustically in the changing formant frequencies just before and after the moment of occlusion and in the spectrum of the release burst. Thus, an intervocalic context would seem to be the best one for correct identification by ear. A final released stop has an incipient vowel, voiceless after a voiceless closure, and so is close to the intervocalic case. There are languages, such as French and German, with a high frequency of stop releases. A final unreleased stop has only the articulatory movement to the closure, so it is not surprising that reports in the literature [e.g. Householder, 1956; Malécot, 1958], which concentrated on English, present evidence of greater intelligibility of stops with releases. Other studies [Halle et al., 1957; Delattre, 1958; Wang, 1959] emphasize the perceptual relevance of both formant transitions and release-bursts. We wonder then whether in Thai, with its unreleased final stops, there is some kind of articulatory compensation that provides useful auditory enhancement of the closing gesture. By the same token, it remains to be demonstrated that Thai listeners have no difficulty in recognizing the unreleased stops in their own language when they are presented in isolated words.

We have begun addressing the topic by having both native Thai speakers and native English speakers identify randomized lists of Thai monosyllables comprising sets within which the members were differentiated by the four final stops of Thai. We hypothesized that (1) Thai listeners would achieve high scores in the identification of the stops, (2) phonetically naive English listeners would do much less well, with especially poor results for the final glottal stop, and (3) phonetically trained English listeners would fall between the other two groups with scores approaching those of the native Thai.

Such an outcome would prompt, in a sequel to this work, a test of the aforementioned gestural hypothesis through instrumental examinations of articulation. That is, our English listeners, after all, are quite used to hearing unreleased stops, as well as released ones, in English, so, we reasoned, if the expected perceptual difference between the two groups of listeners is indeed obtained, there could be an articulatory difference between the two languages in the production of final stops. Thus, it may be that information on place of articulation is enhanced in the formant transitions of Thai final oral stops by characteristics of the movement toward closure. The observation that at least for the apical stops the Thai place of articulation is dentialveolar [Harris, in press], slightly forward of the English alveolar closure, is suggestive of a difference in trajectory. If the Thai listeners did significantly better with the unreleased final oral stops of Thai, this would lend plausibility to the gestural hypothesis by way of the argument that English speakers could not take advantage of the additional information in Thai because of their inexperience with these foreign articulatory enhancements of the sounds. (Of course, it follows that this plausibility would have to be validated by detailed examination of the articulatory gestures themselves.) Poor perception of the glottal stop by English speakers, however, would also have to be explained by its nondistinctive function in the English speech of our subjects.

Procedure

Certain constraints of Thai phonology limited our choice of syllables for the perception tests. As stated in the 'Introduction', the vowel in each syllable had to be short to allow for a possible final glottal stop. Also, Thai is a tone language. In Standard Thai, the official language of Thailand and the variety used in this study, there are five phonemic tones. In short syllables ending in a stop, however, only the high tone /´/ or the low tone /`/ can occur. This is the canonical form of such syllables; there are a few exceptions in the lexicon. The use of nonsense syllables to make the stimuli equally meaningless to both language groups was not really an option, because too many of the phonologically allowable monosyllables are in fact words in the Thai lexicon.

Table 1. Thai words used as stimuli

sàp	to chop, mince	sàt	animal
sàk	to tattoo	sàʔ	to pile (things for a fence)
láp	to sharpen, hone	lát	to take a shortcut
lák	to steal	láʔ	to abandon
lép	fingernail	lét	to leak out
lék	small	léʔ	soggy, soft
bùp	to grind into pieces	bùt	a child
bùk	to force one's way into	bùʔ	to line (the inside of a bag)
dàp	to extinguish	dàt	to bend
dàk	to trap	dàʔ	(to attack) indiscriminately
júp	to collapse	jút	battle
júk	era	júʔ	to urge, to incite

Stimuli

The Thai words chosen for recording are given in table 1 in transcription with glosses. The list contains four sets of monosyllabic words, one set for each of the four final stops. Each set contains six words, three with the high tone, and three with the low tone. Given the aforementioned phonological constraints and our desire to have several four-member sets of well-known Thai words highly frequent in occurrence, the three short vowels /e a u/ and five initial consonants /b d l j s/ were unevenly distributed; however, the syllables in which they occurred were minimally contrasted with other such syllables across the sets by the final stops. That is, as seen in the groupings of table 1, for every combination of initial consonant plus vowel plus tone there were four words minimally differentiated by the final stop consonants.

A highly educated Thai professional woman (speaker 1), a native speaker of Thai but also fluent in English, who was available at the time in New Haven, Conn., recorded several readings of the list with a pause after each word, yielding, in effect, one-word utterances. This was done in a sound-insulated IAC booth at Haskins Laboratories with a Sennheiser MKH415T directional microphone and a Panasonic SV-3700 Professional Digital Audio Tape Deck at a sampling rate of 44.1 kHz. With clarity and general sameness of sound levels as our criteria, as judged by ear, we chose two tokens of each word and saved them as Sound Designer II© files at 22,050 Hz for the PsyScope© computer program. Two tokens of each of the 24 words in table 1 yielded a total of 48 sdtimuli to be labeled by our listeners in each sessions.

The PsyScope program enabled us to design a listening test that could be played in a new random order at each session for each subject. (As will be explained, unanticipated exigencies of the situation in Thailand mandated a somewhat different procedure for administering the listening tests there.) Each stimulus was to appear in a paired presentation. That is, the subject heard the same token of each word twice before giving a response. A description of the subjects, method, and results will be presented separately for each group of subjects.

Data Reduction

Our three groups of test subjects – naive American, Thai, skilled American – differed in both membership size and number of responses to the stimuli; therefore, we have presented the responses as percentages and processed them as percentages in our analyses of variance and t tests to facilitate comparisons across the groups. This use of percentages rather than the raw data might be seen as a source of trouble for statistical analysis, especially for cells with extreme values. Although it is a standard remedy to perform the arc sine transform on all percentages, it should be noted that for our data, as will be seen, the significance values are very strong. It is well known that ANOVA is robust with regard to heterogeneity of variance in the cells, so we have not proceeded further [Ott, 1988, p. 420].

Although the response data are presented as percentages in confusion matrices (tables 2–4), they are readily convertible to raw numbers, as the total number of responses to each stimulus is given in every table heading. The matrices themselves receive no statistical treatment, for we have been unable to find an appropriate technique. One that comes to mind, χ^2, is inappropriate because each subject contributes to more than one cell. In our discussions of the matrices as such, then, we treat the data informally to give the reader some information about what kind of errors were made.

Experiment I: Naive English Speakers

Subjects

Our group of phonetically naive native speakers of northeastern American English, specifically that of southern New England, comprised 19 undergraduate students of introductory psychology at the University of Connecticut in Storrs, Conn., who reported no hearing loss. Except for rudimentary instruction in the pronunciation of one or another Western foreign language, they had had no serious exposure to training in phonetics, and they had had no experience with the Thai language.

Method

Under control of the PsyScope program on a Power Macintosh computer, for each session the stimuli were randomized anew and played through high-quality headphones. The 19 naive American listeners were tested on-line one person at a time in a quiet laboratory in the Psychology Building at the University of Connecticut. They were told that they would hear monosyllabic words in a foreign language ending in sounds that they could identify by striking the keys 'P', 'T', or 'K', on the computer keyboard, or a fourth foreign speech sound that they should call 'other' by striking the 'O' key. It might be objected here that since glottal stops are widely observed in English, English-speaking subjects could easily be taught to use the symbol /ʔ/ or some other mark upon hearing a final glottal stop in the experiment. We considered this to be highly unlikely for the phonetically naive American subjects used by us for whom postvocalic glottal stop – or indeed prevocalic glottal stop – is not phonologically relevant. In their speech glottal stop can occur at the onset of a stressed vowel at the beginning of an utterance and perhaps sporadically in some contexts under weak stress where /t/ or a flap is expected in standard American English. (There are indeed certain varieties of English, e.g. Scots English and Cockney, in which glottal stop seemingly has to be treated as a phoneme and included in the canonical forms of many words.) We felt strongly that if the naive English speakers could label final /ptk/ with great confidence, they would only have to decide that certain stimuli did not end in those oral stops but rather in something foreign for which they had no spelling, something that they could label 'other'. If, however, they tended to identify the glottal stop with other response categories, that too would be informative. This was, then, a four-way forced choice test.

After a practice session of six items in paired presentations to accustom the listeners to the voice of the speaker and the use of the keyboard, the subjects heard each of the 48 stimuli twice through headphones before having to respond. Only after a response was given did the paired presentation of the next stimulus come. Thus the listener controlled the length of the interval in which he or she had to choose a response. That is, both for the practice items and the test proper the presentation was self-paced. The responses were put into a data file by the program. Over a period of 3 days, each subject was exposed to four randomizations of the stimuli recorded by speaker 1.

Results

In table 2 we have a confusion matrix for the responses of the phonetically naive American English listeners to the utterances of Thai speaker 1. A one-way repeated measures analysis of variance of the percentages correct [$F(3,54) = 219.74$, $p < 0.0001$] shows the effect of place of articulation of the stop consonants to be highly significant. This is clearly seen in table 2 in the cells showing only 11% of the responses to the final glottal stops correctly labeled 'other' and 71% of them called 'T'. In addition, the mere 71% correct for /k/ is mainly caused by confusion with /p/ and /t/.

Table 2. Confusion matrix of 19 naive American English listeners' responses in percentages to Thai speaker 1's words with final stops

Responses	Stimuli p	t	k	ʔ
p	*94.7*	6.0	13.5	7.7
t	1.9	*90.7*	10.8	70.8
k	1.6	1.6	*71.4*	10.2
Other	1.8	1.6	4.3	*11.3*

The percentages correct are in italics; n = 912 responses to each consonant.

Table 3. Confusion matrices of 30 Thai listeners' responses in percentages to the 2 Thai speakers' words with final stops

Responses	Stimuli p	t	k	ʔ
Speaker 1				
p	*87.9*	2.4	7.4	0.4
t	3.3	*96.7*	0.6	0.4
k	7.8	0.6	*92.0*	1.0
ʔ	1.1	0.4	0.1	*98.3*
Speaker 2				
p	*82.2*	0.4	1.3	0.1
t	9.2	*98.6*	0.4	0
k	8.3	0.9	*98.0*	0.1
ʔ	0.2	0.1	0.4	*99.9*

The percentages correct are in italics; n = 720 responses to each consonant.

Experiment II: Native Thai Speakers

Stimuli

The utterances of speaker 1, recorded in the United States, formed one set of stimuli. Although we had no reason to doubt the quality of these recordings, we deemed it prudent in Thailand, with no shortage of speakers, to obtain a second set for additional testing. If idiosyncratic differences between the productions of the 2 speakers had significant perceptual effects for our Thai subjects, it would presumably be necessary to obtain data for both speakers' utterances from American subjects.

Another highly educated Thai woman, a member of the academic community, served as speaker 2. A native speaker of Thai, she was also fluent in English. She too recorded several readings of the list of monosyllabic words in table 1 with a pause after each word, yielding, in effect, one-word utterances. This was done in a sound-insulated studio of the Linguistic Research Unit of the Faculty of Arts, Chulalongkorn University, Bangkok. We used a dynamic microphone that sent the signal through the amplifier of a Sony WM-D6C Professional tape recorder to the Apple Sound Input Device of a Power Macintosh laptop computer for sampling at the rate of 22,050 Hz. Once again, with clarity and general sameness of sound level as our criteria, as judged by ear, we chose two tokens of each word and saved them as Sound Designer II files for the PsyScope program. Two tokens of each of the 24 words in table 1 yielded a total of 48 stimuli to be labeled by our listeners in each test.

Subjects

The Thai subjects were 30 first-year undergraduate students in the Faculty of Arts, Chulalongkorn University in Bangkok, all of them native speakers of Standard Thai who reported no hearing loss and had had no special training in phonetics. Having begun their study of English in either the 3rd or 5th year of their elementary education, they all had some command of the language. A number of them also had had a few years' study of a second foreign language.

Method

Our original plan was to test the Thai subjects one at a time on-line, as we had done with the American subjects. Unfortunately, conflicts with other projects and the exigencies of the academic schedule necessitated administering the tests not too long before the time of the final examinations for the academic year. Our solution was to prerecord four randomizations of the test from the PsyScope program with the utterances of speaker 1 and four with those of speaker 2 on magnetic tape and administer the tests to all 30 subjects through headphones in two 1-hour sessions in a new language laboratory at Chulalongkorn University. In a pilot test comparing the two methods we could not convince ourselves that either one was easier for listeners. For us, the experimenters, the use of the language laboratory had the time-consuming disadvantage of requiring the manual transfer of all responses to computer files.

All instructions were given in Thai by the second author (K.T.). Here too a modification of the procedure was unavoidable. For our Thai listeners, of course, all the monosyllables of table 1 were words of their language. Without keys to press, the subjects would have to write their responses. Certain orthographic irregularities in the writing of word endings made it impractical to have them write spellings for all the final consonants. Rather, for each randomization the subject had an answer sheet with the full set of 24 words printed in Thai script at the top. The task was to listen to each paired presentation, decide within an interval of 3 s what word of the set had been said, and write that word in Thai script in the appropriate numbered space on the answer sheet. (During the playing of the six practice items at the beginning, they wrote nothing but looked at the sample correct answers that had been written beforehand in the six spaces by us.)

Results

The responses of the Thai listeners to the utterances of the 2 Thai speakers are given as percentages in confusion matrices in table 3. The diagonal line running through the cells from the upper left corner to the lower right corner gives the percentages of correct responses. A two-factor repeated measures analysis of variance of the latter reveals [$F(1,29) = 1.65$, $p = 0.2092$, n.s.] that the effect of the speaker is not significant. The effect of consonant is highly significant [$F(3,29) = 83.14$, $p < 0.0001$], as is the interaction of speaker and consonant [$F(3,87) = 16.01$, $p < 0.0001$]. The latter effect can be seen in table 3 in the superior response to /p/ and inferior response to /k/ of speaker 1 compared with those of speaker 2.

Experiment III: Skilled English Speakers

Subjects

We also obtained the responses of 7 skilled listeners to the productions of speaker 1. These 7 people, all native speakers of American English, have had some combination of much experience in speech research and advanced training in phonetics. Six of them reported no hearing loss. One of them, somewhat older than the others, had suffered some loss, especially in the higher frequencies, through presbycusis; however, his long experience in phonetics, both instrumental and practical, seemed to warrant keeping him in the group. Indeed, given a spectrum limited to 11,000 Hz (one half the sampling rate) and the need presumably to detect formant transitions well below that limit, we thought it worth the risk.

Method

As they were available for only one test session, each person provided just 12 responses to each stop consonant. They were tested on-line one person at a time in an insulated booth at Haskins Laboratories in New Haven, Conn. We wondered whether these phonetically sophisticated individuals would, despite their native English, respond in much the same way as the Thai subjects. Of course, the Thai subjects responded by writing the words in Thai script; the skilled listeners did so by striking the keys for 'P', 'T', 'K', or lower-case 'ʔ', i.e. '/', to serve as the symbol for glottal stop.

Results

The responses of our 7 skilled listeners are given in table 4. A one-way repeated measures analysis of variance of the percentages correct [$F(3,18) = 5.25$, $p < 0.01$] shows the effect of place of articulation to be significant. This must be largely attributed to the low percentage correct for /k/ in table 4. Our older listener with some high-

Table 4. Confusion matrix of 7 skilled American listeners' responses in percentages to Thai speaker 1's words with final stops

Responses	Stimuli			
	p	t	k	ʔ
p	*91.7*	3.6	2.4	1.2
t	2.4	*96.4*	7.1	9.5
k	3.6	0	*76.2*	4.8
ʔ	2.4	0	14.3	*84.5*

The percentages correct are in italics; n = 84 responses to each consonant.

frequency hearing loss had, in fact, the fifth highest overall correct responses among the 7 subjects.

Comparisons of the Three Groups

Taking the percentages correct of the upper panel of table 3 and those of tables 2 and 4, we did an analysis of variance of the responses of the Thai speakers together with those of the two groups of English speakers, naive and skilled, to the utterances on which all three groups had been tested, those of speaker 1. The effect of the group [$F(2,53) = 127.41$, $p < 0.0001$] is highly significant, as are the effects of consonants [$F(3,159) = 108.50$, $p < 0.0001$] and consonants by groups [$F(6,159) = 154.44$, $p < 0.0001$]. Again, we find the main contributing factor to be the great differences in the responses to final glottal stop. Another contributing factor is the considerably poorer identification of /k/ in table 2. The overall interactions between the three groups of listeners and the cell means of the percentages of correct responses to the four places of articulation are shown in the plot of figure 1.

For a more refined understanding of figure 1, we have done an unpaired two-tailed t test of the percentages correct for the three groups of listeners. Given that we are making three pairwise contrasts among the three groups, we have required a Bonferroni-corrected significance value of $p = 0.01$. The results are shown in table 5 for the responses to each of the four stop consonants. It should be borne in mind that the naive American subjects' 'correct' responses to the glottal stop were labeled 'other'.

Discussion

By and large, except for one curious discrepancy, the phonetically naive American subjects perceived the final stops less accurately than the Thai in experiment I. The discrepancy is with the category-final /p/. In table 2 we see that the American listeners achieved 95% accuracy, while table 3 reveals only 88% for the Thai listeners' responses to the same source, speaker 1. (Note, by the way, that for speaker 2 the Thai listeners' accuracy was even lower at 82%.) The main factor giving rise to the superi-

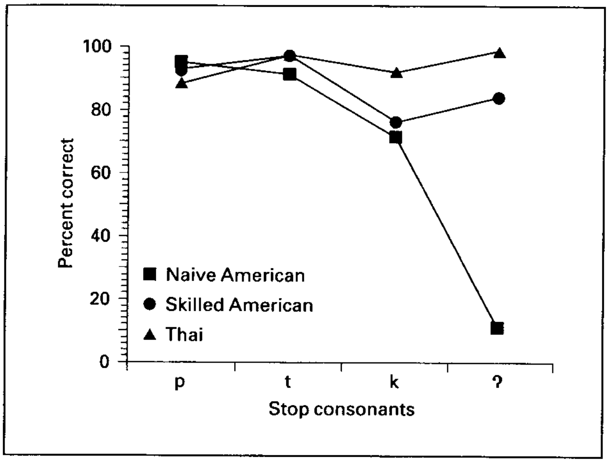

Fig. 1. Plot of the percentages correct of the three groups of listeners in the perception of the Thai final stop consonants uttered by speaker 1.

Table 5. Unpaired t test of the percentage correct responses of the three groups of subjects to the stimuli of speaker 1: Thai, naive American, and skilled American

C	Thai/naive American				Thai/skilled American				Naive American/skilled American			
	MD	d.f.	t	p	MD	d.f.	t	p	MD	d.f.	t	p
p	–6.8	47	–3.08	0.004	–3.8	35	–1.12	0.27, n.s.	3.0	24	0.95	0.35, n.s.
t	6.0	47	3.19	0.003	2	35	0.12	0.91, n.s.	–5.8	24	–1.58	0.13, n.s.
k	20.6	47	5.66	<0.0001	15.8	35	4.10	0.0002	–4.8	24	–0.56	0.56, n.s.
ʔ	87.0	47	37.12	<0.0001	15.0	35	8.72	<0.0001	–72.0	24	–13.99	<0.0001

C = Consonant; MD = mean difference; d.f. = degrees of freedom.

ority of the scores of the Thai subjects is, of course, the response to the category glottal stop: 98 and 100% for the Thai listeners and 11% for the Americans. The American listeners did not, however, merely show poor accuracy by assigning the tokens of glottal stop to the various response categories at random. Rather, as seen in table 2, they assigned the bulk of them, 71%, to the category /t/.

The high percentages correct in experiment II are certainly not surprising for listeners who are native speakers of Thai. Of course, we must suppose that with excellent stimuli and a fully homogeneous group of highly motivated and completely attentive Thai subjects, 100% would have been achieved for each final stop. What surprised us is the lower level of accuracy for the final labial stops, 88% for speaker 1 and 82% for speaker 2. The matter will be discussed further.

In the ideal case, highly trained and much experienced practical phoneticians might be expected to do as well as the native speakers of Thai, if not better. (Indeed, the first author of this paper did do better than the Thai, especially for those cells in table 3 below 98%, but his extensive experience with the language ruled him out as a subject in experiment III.) Not all of our 7 skilled listeners were people with intensive training in the recognition, mimicry, and transcription of a wide range of speech sounds. Still, their overall labeling behavior, as seen in tables 4 and 5 and figure 1, came rather close to that of the Thai.

Our tentative interpretation of the data is that it is not unreasonable to find in them some support for our hypothesis of the perceptual relevance of a cross-language difference in the articulatory closing gestures. For the two stops requiring lingual closure, /t/ and /k/, the Thai listeners did better than the naive Americans, especially with /k/. Although the difference is smaller for the /t/ stimulus itself, it should be noted that the English speakers confused the glottal stop with /t/ 71% of the time. A possible explanation is that the gesture for the true Thai /t/ is different enough from their own English /t/ to have contributed to their confusion with it of /ʔ/, a speech sound with no phonological significance for them. If, as has been claimed [Harris, 1987, in press], voicing is cut off in final Thai stops by glottal closure at the moment of oral closure, this could contribute to audible differences between the final unreleased stops of the two languages. Offhand, it is hard to see how the presence of a simultaneous glottal closure could help differentiate places of articulation, unless the perceptual effect of its presence is somehow more salient for one or another place.

If the foregoing is plausible, how then do we account for the paradoxical response to the labial stop /p/? Why did the English speakers not have a lower score for it? We would suggest that a labial closing gesture for a voiceless stop may simply be essentially the same across these two languages. Even if correct, that does not explain the evident superiority of the English speakers in their responses to the Thai /p/. Here we can only speculate. The American listeners took the tests one person at a time with no one present after the practice session. Even if, let us say, they customarily used the same labial closing gesture as did the Thai listeners, they may have done somewhat better because of better listening conditions and, given their standing as students of psychology, greater attention to the task. The Thai subjects, as required by the circumstances, sat in a language laboratory where the partial shielding around each seat did not keep them from seeing each other and chatting with their neighbors between tests. Occasional loud sounds outside the building could be heard. Thus, although the playback system and the headphones were excellent, there were some random distractions. If this speculation is plausible, it is all the more impressive that the Thai students did somewhat better with the lingual stops. The glottal stop, as we have suggested, is a special case. (It must be understood that the foregoing speculations was motivated by the anomalous results with /p/. Our trial runs certainly failed to reveal anything unduly disturbing or distracting about the setting of the language laboratory.)

With regard to the difference between the two groups in responding to Thai final /p/, it may be of some interest to look at a parallel study by Lisker [1999] with American English stimuli and American English listeners. His first set of data, previously unpublished, were obtained many years ago by playing randomized stimuli from tape recordings to a group of subjects sitting in a test room, much as we have done with our Thai subjects. For the recent resumption of his project, the subjects were tested one at a time, alone in the test room, under the control of the PsyScope program. We have compared our data with a subset of his, namely those for monosyllabic English words ending in /p/ with the three monophthongal vowels /ɛ a ʊ/, which are closest phonetically to the three Thai vowels used by us.

Lisker reports a 'markedly higher' level of intelligibility in the later tests with an overall improvement of about 15%. Specifically for the identification of final unreleased /p/ in the old work, which yielded generally lower scores than the stops /t k/, the overall accuracy for /p/ after the three vowels was 71.3%. In his new work, however, Lisker reports an average of 98.3% accuracy for final /p/ after the same three vowels. If, despite some uncontrolled factors, especially the use of a new speaker in the later experiment, we assume a reasonable sampling of speakers and listeners for both of his experiments, it is at least interesting to note the differences in results. Perhaps they lend some plausibility to our speculation concerning listening conditions and our analogous results.

Our plan for the future is to do a comparative study of the closing gestures for final stop consonants by Thai and English speakers with our magnetometer. Only the labial and lingual stops will be examined, of course, since the Thai glottal stop has no systematic counterpart in the variety of English under consideration. Given the perceptual data of the present study, our gestural hypothesis leads us to expect to find small but significant differences in the movement trajectories of the articulators between Thai and English for the alveolar and velar stops. If this is confirmed, it will still be necessary to go beyond such a correlation and demonstrate its perceptual relevance. One approach would be to do a thorough acoustic analysis of the formant transitions resulting from the closing gestures in the hope of finding differences that in experiments with a terminal-analog speech synthesizer would affect the identifiability of unreleased final stops for Thai listeners. Another approach would be to capture the gestural differences with the parameters of our articulatory synthesizer and test the acoustic consequences on Thai listeners.

If indeed a difference between the closing gestures of the two languages is found, 'Thai' synthetic stimuli should be better identified by Thai listeners than 'English' stimuli. Such an outcome would be compatible with the gestural hypothesis.

Acknowledgment

This research was supported by NIH grant DC-02717 to Haskins Laboratories, New Haven, Conn., USA. The cost of the first author's travel to Thailand and back was covered by grants from the University of Connecticut Research Foundation. He stayed in Thailand from 4 December 1997 until 4 March 1998 to do this work and other research as a guest of the Faculty of Arts and its Department of Linguistics at Chulalongkorn University, Bangkok. He is grateful for the usual warm reception and gracious help given to him by the Head of the Department, Sudaporn Luksaneeyanawin, and her colleagues. We thank Stephen Frost for his patient guidance into the intricacies of the PsyScope program and Leonard Katz for his advice on statistics. We are grateful to Michele Sancier for her careful running of our tests on the naive American subjects while both authors were in Thailand. Finally, we acknowledge the value of the stimulating response to an earlier version of this paper by Molly Mack and an anonymous reviewer.

References

Abramson, A.S.: The vowels and tones of Standard Thai: acoustical measurements and experiments. Indiana University Research Center in Anthropology, Folklore and Linguistic, Publ. 20 (Bloomington 1962).

Abramson, A.S.: Word-final stops in Thai; in Harris, Noss, Tai phonetics and phonology, pp. 1–7 (Central Institute of English Language, Bangkok 1972).

Crystal, T.H.; House, A.S.: Segmental durations in connected-speech signals: current results. J. acoust. Soc. Am. *83:* 1553–1585 (1988).

Delattre, P.: Unreleased velar plosives after back-rounded vowels (letter). J. acoust. Soc. Am. *30:* 581–582 (1958).

Diehl, R.L.; Kluender, K.R.: On the objects of speech perception. Ecol. Psychol. *1:* 121–144 (1989).

Fowler, C.A.: An event approach to the study of speech perception from a direct-realist perspective. J. Phonetics *14:* 3–28 (1986).
Fowler, C.A.: Speech perception: direct realist theory; in Asher, Encyclopedia of language and linguistics, pp. 4199–4203 (Pergamon, New York 1994).
Halle, M.; Hughes, G.W.; Radley, J.-P.A.: Acoustic properties of stop consonants. J. acoust. Soc. Am. *29:* 107–116 (1957).
Harris, J.G.: Linguistic phonetic notes (Chulalongkorn University, Bangkok 1987).
Harris, J.G.: States of the glottis of Thai voiceless stops and affricates; in Tingsabadh, Abramson, Essays in Tai linguistics (Chulalongkorn University Press, Bangkok, in press).
Householder, F.W., Jr.: Unreleased PTK in American English; in Halle, Lunt, McLean, van Schooneveld, For Roman Jakobson, pp. 235–244 (Mouton, The Hague 1956).
Liberman, A.M.; Mattingly, I.G.: The motor theory of speech perception revised. Cognition *21:* 1–36 (1985).
Liberman, A.; Mattingly, I.: A specialization for speech perception. Science *243:* 489–494 (1989).
Lisker, L.: Perceiving final stops without release: effects of preceding monophthongs vs. non-monophthongs. Phonetica *56:* 44–55 (1999).
Malécot, A.: The role of releases in the identification of released final stops: a series of tape-cutting experiments. Language *34:* 370–380 (1958).
Noss, R.B.: Thai reference grammar (Foreign Service Institute, Washington 1964).
Ott, L.: An introduction to statistical methods and data analysis (PWS-Kent Publishing Co., Boston 1988).
Tingsabadh, K.; Abramson, A.S.: Thai. J. int. phonet. Ass. *23:* 24–28 (1993).
Wang, W.S.-Y.: Transition and release as perceptual cues for final plosives. J. Speech Hear. Res. *2:* 66–73 (1959).