Cross-language perception of word-final stops in Thai and English*

This study examined Australian English (AE) and Thai–English bilingual (TE) speakers’ ability to perceive word-final stops in their native and non-native languages. In the perception experiment, the TE listeners were able to discriminate stop contrasts differing only in place of articulation (/p/–/t/, /p/–/k/, /t/–/k/) in both English and Thai accurately, but the AE listeners’ discrimination was accurate only for English. The listeners’ discrimination accuracy was differentially influenced by the type of stop contrast they heard. The Thai /p/–/t/ contrast was most discriminable for both groups of listeners, in particular, the AE listeners. Acoustic analyses of the Thai stimuli presented in the perception experiment were conducted in order to search for cues that led to different response patterns for the AE and TE listeners. There was a clear effect of the final stop on the formant trajectories of /a/ and /u/, suggesting that these acoustic differences may be audible to the listeners. The results provide further evidence that first language (L1) transfer alone is insufficient to account for listeners’ response patterns in cross-language speech perception and that it is necessary to take into account phonetic realization of sounds and/or the amount of acoustic information contained in the speech signal to predict accuracy with which sound contrasts are discriminated.

A number of studies of cross-language or second language (L2) speech perception have focused on the identification and/or discrimination of non-native contrasts that are absent in listeners’ first language (L1) such as English /l/ and /r/ for native Japanese listeners (e.g. MacKain, Best and Strange, 1981; Sheldon and Strange, 1982; Guion, Flege, Akahane-Yamada and Pruitt, 2000; Aoyama, Flege, Guion, Akahane-Yamada and Yamada, 2004) and identified several factors that influenced L2 learners’ success in identifying or discriminating English approximants.1 These factors included the liquid sounds’ position in a word, type of speech stimuli (natural vs. synthetic), lexical familiarity, previous linguistic experience and degree of perceived dissimilarity among others. Best and Strange (1992) showed that cross-language discrimination accuracy depended on the cross-linguistic phonemic assimilation patterns and listeners’ experience with spoken English.

There is ample evidence that native and L2 speakers of English differ in their perception of the voicing and/or place contrast in word-final consonants (e.g. Flege and Wang, 1989; Crowther and Mann, 1992; Flege and Liu, 2001; Aoyama, 2003). A recent study (Aoyama, 2003) examined the perception of syllable-initial and syllable-final English nasals by Korean and Japanese L2 learners. The Japanese listeners found it difficult to discriminate the /n/–/ŋ/ contrast in word-final position, but not the /m/–/n/ or /m/–/ŋ/ contrast although none of these nasal contrasts occur word-finally in their L1. It was observed that the English /m/ was assimilated to a single Japanese category /mu/, but the English /n/ and /ŋ/ were assimilated to multiple Japanese categories, implying perceptual difficulties.

These results appear to be consistent with predictions generated by models of cross-language speech perception such as the Perceptual Assimilation Model (PAM) (Best, 1993, 1995). According to the PAM, instances of contrastive L2 categories that are identified as instances of a single L1 category (e.g. English /l/–/r/ for Japanese learners) will be relatively difficult to discriminate whereas instances of contrastive L2 sounds that are mapped onto two different L1 sounds (e.g. English /w/–/ʃ/ for Japanese learners) will be discriminated more accurately.

Even when native and non-native languages share a phonetic contrast at the abstract phonological level, listeners may fail to show accurate perception in the non-native stimuli. In other words, positive transfer from the L1 may not be reflected in listeners’ performance. An example of this was demonstrated by Hallé, Best and Levitt (1999) for native French listeners’ perception of American English approximants /w j r l/. These phonemes occur in French with varying degrees of
similarity to the target sounds. If L1-to-L2 mappings at the traditional phonological level predict cross-linguistic perception patterns, French listeners would not have difficulties with these English sounds. However, it was found that the French listeners had some perceptual difficulties with the English /r/. This finding was attributed to marked articulatory-phonetic differences between the English and French /r/ (i.e. phonetically realized as a central approximant in English and a uvular fricative in French). Of the three contrasts tested (/w/-/j/, /r/-/l/, /w/-/r/), the French listeners had most difficulty with /w/-/r/ and tended to hear the English /r/ as /w/-like.

The present study examines the cross-language perception of final English and Thai stops by two groups of listeners: native Thai–English bilinguals (TE) and Australian English (AE) speakers living in Australia. Both English and Thai permit /p t k/ in final position. This leads to an expectation that AE listeners would not have difficulty discriminating Thai stop contrasts. However, Abramson and Tingsabadh (1999) showed that American English speakers identified the Thai final stops less accurately than native Thai speakers.

Since only limited counterexamples of expected L1 positive transfer have been reported in the literature, it is considered necessary to examine different language comparisons and syllable position before any generalization could be made. The present study sought to verify if positive transfer at the phonological level might be observed in final position as well as in initial position. To be more specific, this study examined the perception of native and non-native stop place contrasts in word-final position which are functional (i.e. phonemic) in both Thai and English, but are phonetically realized differently as described below.

Word-final stops are invariably unreleased in Thai and variably released in English. Previous research identified several factors that influence the frequency of occurrence of final release bursts in English. These include the vowel type that precedes the final stop (Parker and Walsh, 1981; Lisker, 1999), gender of the talker (Byrd, 1992, 1993, 1994), place of articulation (Crystal and House, 1988; Byrd, 1993), dialect (Byrd, 1992), speaking style (Picheny, Durlach and Braida, 1985, 1986; Bond and Moore, 1994) and the position of the stop within the utterance (Halle, Hughes and Radley, 1957). Although the presence or absence of a release does not signal a change in meaning in English or in any other languages, released stops are known to be more intelligible than unreleased stops (Householder, 1956; Malécot, 1958; Wang, 1959). Word-final stops in Thai, on the other hand, are always unreleased (Tingsabadh and Abramson, 1993) as in some other Asian languages such as Cantonesan, Korean, Taiwanese and Vietnamese.

With respect to discriminability of sound contrasts that occur in different syllable position, it is not entirely clear whether consonants are inherently easier to discriminate in CV than VC syllables. Using synthetic speech stimuli, Jusczyk (1977) showed that two-month-old infants were sensitive to place of articulation differences in CV and VC syllable pairs (target sounds underlined) ending with released final stops and that they did not have any more difficulty with stop contrasts in final position than in initial position. Thus, there may be no inherent asymmetry in discriminability for CV and VC syllables. However, it is conceivable that asymmetry exists and listeners need prior linguistic experience to be able to perceive stop contrasts accurately. Perhaps, already at two months, infants in an English-speaking environment were learning to discriminate contrasts in both syllable positions that are legitimate in their L1.

The primary aim of this study was to compare the perception of native Thai bilinguals and that of Australian English speakers when they heard final stops contrasting only in place of articulation (i.e. /p t k/) in their native and non-native languages. Given that both English and Thai share the same place of articulation for final stops (i.e. /p t k/), a second aim was to determine if clear L1-to-L2 mappings might guarantee a positive transfer and result in accurate perceptual discrimination of stop contrasts regardless of how such contrasts are phonetically realized.

**Perception experiment**

The purpose of this experiment was to assess the discrimination of English and Thai stop place contrasts in final position by AE and TE listeners.

**Method**

**Stimuli**

Tables 1 and 2 show English and Thai words collected to be used as stimuli in this experiment. All test words were monosyllabic /CVC/ words ending in /p t k/. Stimuli were selected from the available tokens produced by three speakers per language. The speakers read the test words in isolation (all real words in English or Thai) in their L1s in the MARCS Auditory Laboratories recording studio at the University of Western Sydney, Australia. Test words were presented visually to each speaker in randomized

---

2 Six pairs of synthetic speech stimuli were tested: “bad”–“bag”, “bag”–“bam”, “ad”–“ag” and “dab”–“gab”, “bad”–“bad” and “bag”–“bag”.

3 Although not stated explicitly, the infants presumably come from an English-speaking background.

4 As described in the subsequent section, the objective of CDT is to assess discrimination accuracy for a contrast (e.g. /p/-/k/) in an oddity discrimination test. As such, even where only minimal pairs, but not minimal triplets, are available (e.g. “meet”–“meek”, “tape”–“take” in Table 1), the words could still be used to test the /N/-/k/ or /p/-/k/ contrast in change trials or three instances of “meet” by different speakers could be presented in a no-change trial.
The recorded speech materials were digitized at 44.1 kHz. Pronounced. Ten unreleased tokens included three release burst although speakers were not given specific of English final stops were produced with an audible tones were collected so that the discrimination of tokens words that had either high or low tones were written orders on the computer screen one word at a time. Thai words which differed solely in lexical tones could be examined in a separate study. In order to prevent the speaker gender from affecting listeners’ responses, male and female voices were not presented together within a given trial, but instead, tokens from three female speakers were used for English stimuli and tokens from three male speakers were used for Thai stimuli.

More than 90% (107 out of 117 unique waveforms) of English final stops were produced with an audible release burst although speakers were not given specific instructions as to how the final stops should be pronounced. Ten unreleased tokens included three instances of /p/, six instances of /t/ and one instance of /k/. The recorded speech materials were digitized at 44.1 kHz using CoolEdit and normalized for peak intensity (50% of the full scale) using the UAB software (Smith, 1997).

### Listeners

Two groups of listeners participated: A group of 26 native Thai speakers (6 male, 20 female) with a mean age of 29.8 years (sd = 10.6, range = 17–57) and a group of 16 AE listeners (3 male, 13 female) with a mean age of 24.0 years (sd = 9.1, range = 18–48). Thai bilinguals’ mean length of residence (LOR) in Australia was 5.3 years (sd = 6.0, range = 0.2–30.3). Their mean age of arrival (AOA) in Australia was 24.6 years (sd = 6.6, range = 14–41). The TE participants were paid for their participation. The AE participants were students at the University of Western Sydney or Macquarie University. Ten AE listeners tested in the MARCS Auditory Laboratories at the University of Western Sydney were enrolled in a first-year Psychology course and six AE listeners tested in the Department of Linguistics at Macquarie University were enrolled in a first or second-year Linguistics course. They received credit points for the respective courses for their participation. None of them had any knowledge of Thai. All participants were tested individually in a single session.

### Table 1. English words recorded for the perception experiment.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>/p/</th>
<th>/t/</th>
<th>/k/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>peep</td>
<td>Pete</td>
<td>peak</td>
</tr>
<tr>
<td>/b/</td>
<td>beep</td>
<td>beat</td>
<td>beak</td>
</tr>
<tr>
<td>/s/</td>
<td>seep</td>
<td>seat</td>
<td>seek</td>
</tr>
<tr>
<td>/ˈɛ/</td>
<td>/ˈæ/</td>
<td>/ˈeɪ/</td>
<td>/ˈeɪ/</td>
</tr>
<tr>
<td>/ʃ/</td>
<td>cheap</td>
<td>cheat</td>
<td>cheek</td>
</tr>
<tr>
<td>/k/</td>
<td>weep</td>
<td>wheat</td>
<td>week</td>
</tr>
<tr>
<td>/u/</td>
<td>sip</td>
<td>sit</td>
<td>sick</td>
</tr>
<tr>
<td>/l/</td>
<td>lip</td>
<td>hit</td>
<td>lick</td>
</tr>
<tr>
<td>/ˈɛ/</td>
<td>/ˈæ/</td>
<td>/ˈeɪ/</td>
<td>/ˈeɪ/</td>
</tr>
<tr>
<td>/o/</td>
<td>rap</td>
<td>rat</td>
<td>rack</td>
</tr>
<tr>
<td>/ˈæ/</td>
<td>/ˈæ/</td>
<td>/ˈeɪ/</td>
<td>/ˈeɪ/</td>
</tr>
<tr>
<td>/ɒ/</td>
<td>cap</td>
<td>cat</td>
<td>back</td>
</tr>
<tr>
<td>/æ/</td>
<td>lap</td>
<td>suit</td>
<td>suit</td>
</tr>
<tr>
<td>/ɒ/</td>
<td>carp</td>
<td>heart</td>
<td>heart</td>
</tr>
<tr>
<td>/ɑː/</td>
<td>sharp</td>
<td>toe</td>
<td>shark</td>
</tr>
</tbody>
</table>
| /r/   | cup| cut| -*
| /ɒ/   | cope| coat| coke|
| /æ/   | * | fate| fake|
| /ɒ/   | * | late| lake|
| /æ/   | * | mate| make|
| /ɑː/  | tape| toe| take|
| /æ/   | * | bite| bike|

Note: The symbol “*” indicates nonwords and they were not recorded for the perception experiment.

### Table 2. Thai words recorded for the perception experiment. Gloss available upon request.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Tone</th>
<th>/p/</th>
<th>/t/</th>
<th>/k/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>low</td>
<td>ติ/</td>
<td>ขิ/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/a/</td>
<td>low</td>
<td>คิ/</td>
<td>ขิ/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/e/</td>
<td>low</td>
<td>คา/</td>
<td>คิ/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/o/</td>
<td>low</td>
<td>หา/</td>
<td>หา/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/ɔ/</td>
<td>low</td>
<td>วา/</td>
<td>วา/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/a/</td>
<td>low</td>
<td>ติ/</td>
<td>ขร/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/ɒ/</td>
<td>low</td>
<td>คิ/</td>
<td>ขร/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/æ/</td>
<td>low</td>
<td>หา/</td>
<td>หา/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/o/</td>
<td>low</td>
<td>หา/</td>
<td>หา/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/ɒ/</td>
<td>low</td>
<td>ว่/</td>
<td>ว่/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/ɛ/</td>
<td>low</td>
<td>คิ/</td>
<td>คิ/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/æ/</td>
<td>low</td>
<td>คิ/</td>
<td>คิ/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/ɑː/</td>
<td>high</td>
<td>คา/</td>
<td>คา/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/ɒ/</td>
<td>high</td>
<td>ตาม/</td>
<td>ตาม/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/o/</td>
<td>high</td>
<td>หา/</td>
<td>หา/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/ɔ/</td>
<td>high</td>
<td>หา/</td>
<td>หา/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/æ/</td>
<td>high</td>
<td>หา/</td>
<td>หา/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/o/</td>
<td>high</td>
<td>หา/</td>
<td>หา/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/ɒ/</td>
<td>high</td>
<td>หา/</td>
<td>หา/</td>
<td>ไม/</td>
</tr>
<tr>
<td>/æ/</td>
<td>high</td>
<td>หา/</td>
<td>หา/</td>
<td>ไม/</td>
</tr>
</tbody>
</table>

Note: The symbol “*” indicates nonwords and they were not recorded for the perception experiment.
lasti ng about 45–60 minutes. All of them reported normal hearing and had no history of language problems.

**Task**

A categorial discrimination test (CDT) employed in previous L2 speech research (e.g. Flege, MacKay and Meador, 1999; Guion, Flege, Akahane-Yamada and Pruitt, 2000; Flege, 2002; Flege and MacKay, 2004) was used. The stimuli (117 and 125 unique waveforms selected from words in Tables 1 and 2 for English and Thai, respectively) were presented in triads via headphones at a self-selected comfortable level using a notebook computer. Some waveforms were used more than once in different trials in combination with different tokens. Each contrast was tested by change and no-change trials. The three stop tokens in all change and no-change trials were spoken by different talkers, and so were always physically, if not phonetically, different. Listeners were asked to choose a word that was different from the other two, if there was any.

The change trials contained an odd item out. For example, a change trial testing the /p/–/t/ contrast might consist of “sip2”–“sit1”–“sip3” (where the subscripts indicate different talkers). In other words, in change trials, listeners would hear minimal pairs differing only in the place of final stop (i.e. the initial consonant, the medial vowel and the lexical tone (in the case of Thai) were kept constant in each trial) produced by three different speakers. The correct response for change trials was the button (“1”, “2” or “3”) indicating the position of the odd item out, which occurred with near-equal frequency in all three possible serial positions. The change trials tested the participants’ ability to respond appropriately to relevant phonetic differences between tokens and distinguish stops drawn from two different categories. The correct response to no-change trials, which contained three different instances of a single category (e.g. /p/1/p/3/p/2 or /t/1/t/1/t/2), was a fourth button marked “NO”. The no-change trials tested the participants’ ability to ignore audible but phonetically irrelevant within-category variation (e.g. in voice quality).

The participants were required to respond to each trial, and were told to guess if uncertain. A trial could be replayed, but responses could not be changed once given. The inter-stimulus interval in all trials was 0.5 s. Two blocks of 58 trials were presented. A different randomization was used for each block. The first ten trials were for practice and were not analyzed. The 48 trials in each block consisted of 36 change trials testing three contrasts (12 trials each for /p/–/t/, /p/–/k/, /t/–/k/) and 12 no-change trials (4 trials each for /p/–/p/, /t/–/t/, /k/–/k/). The English and Thai stimuli were presented in separate blocks and the listeners heard stop contrasts in their own language first and then in the other language. Responses to the change and no-change trials were used to calculate A’ scores (Snodgrass, Levy-Berger and Haydon, 1985), an index of discrimination accuracy. These scores were based on the proportion of ‘hits’ (correct selection of an odd item in change trials) obtained for each contrast and the proportion of ‘false alarms’ (incorrect selection of an odd item in no-change trials). If the proportion of hits (H) equaled the proportion of false alarms (FA), then A’ was set to 0.5. If H exceeded FA, then A’ = 0.5 + ((H-FA)/(1 + H-FA))(4*H)/(1-FA)). However, if FA exceeded H, then A’ = 0.5 - (FA-H)/(1 - FA-H))((4*FA)/(1-H)). For instance, to calculate a score for the /p/–/t/ contrast by a certain listener, the percentage of time an odd item was correctly selected in change (e.g. /p/–/t/–/p/, /t/–/t/–/p/) trials and the percentage of time an odd item was incorrectly selected in no-change (i.e. /p/–/p/–/p/ and /t/–/t/–/t/) trials were used. A score of 1.0 indicated perfect sensitivity, whereas a score of 0.5 or lower indicated a lack of phonetic sensitivity.

**Results**

Figure 1(a) shows mean discrimination scores for Thai stimuli by the TE and AE listeners as a function of contrast types. The overall mean discrimination scores were 0.9 and 0.71 for the TE and AE groups, respectively. Most AE listeners found it much more difficult to discriminate Thai than English stop contrasts. None of the AE listeners’ scores reached the mean score obtained by the TE listeners, suggesting a considerable level of perceptual difficulty for the AE listeners to discriminate Thai final stop contrasts.

A three-way ANOVA with Listener group (AE, TE) as a between-subjects factor and Stimulus language (Thai, English) and Contrast (/p/–/t/, /p/–/k/, /t/–/k/) as within-subjects factors were carried out. The dependent variable was the discrimination score obtained by each listener. All three main effects reached significance [Group: F(1,40) = 11.5, p < 0.01, Stimuli: F(1,40) = 84.5, p < 0.001, Contrast: F(2, 80) = 11.1, p < 0.001] and so did Group × Stimuli and Contrast × Stimuli interactions [G × S: F(1,40) = 72.2, p < 0.001, C × S: F(2, 80) = 13.5, p < 0.001]. A three-way interaction was also significant [F(2, 80) = 7.3, p < 0.01], apparently because of different patterns of two-way interactions for the two stimulus languages as shown in Figures 1(a) and (b).

Two Group × Contrast ANOVAs were therefore conducted for each language separately. For the Thai stimuli, both main effects reached significance [Group: F(1,40) = 48.8, p < 0.001, Contrast: F(2, 80) = 14.7, p < 0.001] and so did a two-way interaction [F(2, 80) = 7.4, p < 0.01]. This was due to a slightly different effect of Contrast for the two listener groups. For the AE listeners, discrimination accuracy was highest for /p/–/t/
Cross-language stop perception

Figure 1. Mean discrimination scores for (a) Thai stimuli and (b) Australian English stimuli by Thai English (TE) and Australian English (AE) listeners. The brackets enclose ± one standard error.

(0.79), lowest for /t/-/k/ (0.64) and intermediate for /p/-/k/ (0.71). While the TE listeners also discriminated /p/-/t/ most accurately (0.92), their scores for /p/-/k/ and /t/-/k/ did not differ (0.89 for both). The simple effect of Group was significant for all three contrasts \(F(1, 40) = 19.0-44.1, p < 0.001\). The advantage of the TE group over the AE group can be clearly seen in Figure 1(a) for all stop contrasts tested.

Figure 1(b) shows discrimination scores for English stimuli by the TE and AE listeners as a function of contrast types. The overall mean discrimination scores were 0.91 and 0.96 for the TE and AE groups, respectively. Most AE listeners were very good at discriminating English stop contrasts. Only one out of 16 listeners obtained a mean score of lower than 0.90.

The effect of Group did not reach significance, but the effect of Contrast did \(F(2, 80) = 7.3, p < 0.01\). Both groups discriminated the /p/-/k/ contrast most accurately. It has been shown that the intelligibility of the final /k/ is particularly susceptible to the presence or absence of a release burst (Cheung, 1986 cited in Matthews and Yip, 1994; Abramson and Tingsabadh, 1999; Lisker, 1999; Tsukada, Birdsong, Mack, Sung, Bialystok and Flege, 2004). The present finding is presumably due, at least in part, to clearly released /k/ tokens in the stimuli. A two-way interaction was not significant, suggesting that all listeners responded in a comparable manner to different stop contrasts.

The TE listeners’ mean discrimination scores for Thai and English hardly differed: 0.90 for Thai and 0.91 for English, respectively, suggesting that they were competent in discriminating final stops in both their L1 and L2. The AE listeners, on the other hand, were much better at discriminating final stops in English than in Thai (0.96 vs. 0.71), possibly because they are less well attuned than the TE listeners to the absence of release in the stimuli (Abramson and Tingsabadh, 1999).
Discussion

The Thai bilinguals were capable of discriminating final stops in both English and Thai whereas the AE listeners showed accurate discrimination only in their L1. As briefly mentioned in the Introduction, all contrasts tested in this experiment were phonemic in both English and Thai. However, there are cross-linguistic phonetic differences in how stops are produced in word-final position — variably released in English and invariably unreleased in Thai. As such, phonetic realizations of final stops are not always equivalent in the two languages. Nevertheless, given the existence of unreleased stops in English, it is somewhat surprising that they did not discriminate unreleased Thai stops more accurately. It may be the case that the AE listeners were disadvantaged, because they had no knowledge of Thai.

Although in this experiment, more than 90% of the entire English stop tokens were released by the AE speakers who read the target words, it is possible that, irrespective of the presence or absence of release burst, place contrasts in English stops are perceptually more salient and discriminable than Thai stop contrasts for any listener. In other words, the TE listeners’ high performance in English may not have resulted from their L2 learning. To confirm or disconfirm this hypothesis, it would be necessary to test native Thai listeners who have minimal exposure to spoken English. If such monolingual Thai listeners still show accurate discrimination of English stop contrasts, it is likely that English stops (at least released stops used in the present experiment) contain more acoustic information than Thai stops do. If, on the other hand, Thai monolinguals’ response patterns mirror those of the AE listeners, i.e. accurate discrimination only in their L1, then the results in this experiment must be a reflection of Thai bilinguals’ L2 perceptual learning.

It is also possible that the AE listeners’ perception of Thai stops may be amenable to training. Unlike Japanese learners of English who do not possess separate phonemic categories for /l/ and /ɹ/, having distinct stop categories in their native language may help AE listeners to make the necessary perceptual adjustments with relatively little training.

Acoustic analysis

The perception experiment above yielded clear between-group differences when listeners responded to the Thai, but not English, stimuli. However, their discrimination accuracy depended on the type of stop contrasts they heard. The Thai /p/--/t/ contrast was most discriminable for both groups of listeners. This section provides acoustic data of the Thai stimuli presented in the discrimination experiment in an attempt to seek possible explanations of the differences in listeners’ response patterns to the three stop contrasts.

Method

Speech materials

All the Thai tokens that were presented to the AE and TE listeners in the perception experiment were examined acoustically. This included a total of 125 unique waveforms representing 64 of 69 words in Table 2. The Thai /CVC/ words included three vowels (/i a u/) which preceded the final stop /p t k/ (see Table 2). The target words were produced by three male native speakers of Thai as described in the perception experiment.

Data processing

The first two formant center frequencies (F1, F2) of the target vowels were automatically tracked in a signal processing package ESPS/Waves (http://www.speech.kth.se/software/). The settings were 12th order linear predictive coding analysis, cosine window, 49-ms frame size, and 5-ms frame shift. The EMU speech database system (http://emu.sourceforge.net/) was used for phonetically labelling the speech segments of interest and formant trajectories were plotted in the R statistical environment (http://cran.r-project.org/). The beginning and end of each vowel token was identified by inspection of wide-band spectrograms and time domain waveforms.

Analysis

F1 and F2 values at vowel offset were measured to assess the effect of the following stop. These values were submitted to Stop (p, t, k) × Vowel type (i, a, u) ANOVAs. Separate analyses were conducted for F1 and F2 values.

Results

Formant trajectories of Thai vowels preceding the three stop places are plotted in Figure 2. Each of the formant trajectories was linearly interpolated, and time was normalized in order to linearly stretch or compress the trajectories such that they were of equal duration. The effect of the final stop on formant trajectories is evident for /a/ and /u/. For these vowels, F2 transition into the final /p/ and /t/ showed distinct directions of movement, suggesting that these acoustic differences may be audible to the listeners.

---

5 Five words (/tʰúp/, /tʰúk/, /hát/, /pák/, /sát/) in Table 2 were not used as stimuli in the perception experiment and were, therefore, not included in the acoustic analysis. This irregularity arose in the stimuli selection process in order to balance as much as possible the number of change trials for the /p/–/t/, /p/–/k/, /t/–/k/ contrasts (12 each) and no-change trials for the /p/–/p/, /t/–/t/, /k/–/k/ contrasts (4 each) that contain the same vowel type.
Cross-language stop perception

Figure 2. Formant trajectories of Thai vowels /i a u/ as a function of the stop place.

For F1, only the effect of Vowel type reached significance \[F(2, 116) = 278.8, p < 0.001\]. As expected, the low vowel /a/ had significantly higher F1 at 709 Hz than the high vowels /i/ (343.8 Hz) and /u/ (319.6 Hz), which did not differ from each other. It can be seen in Figure 2 that the effect of final stop is minimal on F1.

Unlike F1, F2 was considerably influenced by the final stop. The main effects of Vowel type and Stop were highly significant \[Vowel: F(2, 116) = 613.2, p < 0.001; \text{Stop}: F(2, 116) = 41.7, p < 0.001\] and so was a two-way interaction \[F(4, 116) = 57.7, p < 0.001\]. The significant two-way interaction was explored by simple effect tests. As is clearly seen in Figure 2, the simple effect of Stop was significant for /a/ \[F(2, 79) = 132.8, p < 0.001\] and /u/ \[F(2, 28) = 119.9, p < 0.001\], but not /i/. For /a/, F2 was highest before /t/ (1508.8 Hz), lowest before /p/ (1171.7 Hz) and intermediate before /k/ (1404.5 Hz). All comparisons were significantly different. For /u/, F2 was highest before /t/ (1344.8 Hz), lowest before /k/ (568.1 Hz) and intermediate before /p/ (745.8 Hz). Again, all comparisons differed significantly. For both /a/ and /u/, F2 values were substantially contrastive when the vowels occurred before /p/ and /t/. It can therefore be inferred that the listeners were aided by this information in the speech signal when they heard the /p/~/t/ contrast in the perception experiment.

In addition to providing a possible explanation for high scores obtained for the /p/~/t/ contrast, acoustic characteristics shown above generated further predictions for listeners’ discrimination accuracy when the stop contrasts were preceded by different vowels, in particular, /a/ and /u/. Of the total 96 trials (48 × 2 blocks), 8 contained /i/, 62 contained /a/ and 26 contained /u/ as contextual vowels. As for the /a/ vowel, it is anticipated that /t/~/k/ would be difficult whereas for /u/, /p/~/k/ would be difficult. To assess these predictions, discrimination scores were recalculated. For this post-hoc analysis, the percentage of times that each stop contrast following /a/ or /u/ was correctly discriminated was used as a measure of discrimination accuracy. This was because A’ scores which take into account both hits and false alarms for each stop contrast could not always be calculated when the vowel type was included as an additional factor. These new discrimination scores were submitted to a three-way ANOVA with Listener Group (AE, TE) as a between-subjects factor and Contrast (/p/~/t/, /p/~/k/, /t/~/k/) and Vowel (/a/, /u/) as within-subjects factors.

---

Although it would have been desirable to balance the number of trials containing the different vowels, the original design of the present study required a selection of words contrasting in the place of articulation of final stops, and therefore, the vowel type was not considered a primary factor.
Figure 3. Mean percentage of Thai stop contrasts containing (a) /a/ vowel and (b) /u/ vowel correctly discriminated by Thai English (TE) and Australian English (AE) listeners. The brackets enclose ± one standard error.

The ANOVA yielded significant main effects of Group \(F(1, 40) = 82.7, p < 0.001\) and Contrast \(F(2, 80) = 104.8, p < 0.001\), but not Vowel. All three two-way interactions reached significance \(G \times C: F(2, 80) = 7.8, p < 0.001, G \times V: F(1, 40) = 8.9, p < 0.01, C \times V: F(2, 80) = 80.9, p < 0.001\) and so did a three-way interaction \(F(2, 80) = 3.3, p < 0.05\).

Figures 3(a) and (b) show the percentage of times each stop contrast was correctly discriminated by the TE and AE groups for the /a/ and /u/ vowels, respectively. For both listener groups, it can be seen that each stop contrast was differentially influenced by different vowels. Although these results need to be treated with caution, because the number of trials including /a/ and /u/ was not balanced, it appears that the perception and acoustic data do converge. When the context vowel was /a/, both groups discriminated /t/–/k/ less accurately than the other two contrasts and when the context vowel was /u/, they discriminated /p/–/k/ less accurately than the other two contrasts. It is interesting to note in Figure 3(b) that even the TE listeners’ accuracy was substantially lower at 52.5% for /p/–/k/ than for the other contrasts.

Discussion

The perception experiment demonstrated that the AE and TE listeners were able to discriminate all three contrasts (/p/–/t/, /p/–/k/, /t/–/k/) in English accurately. The AE listeners were less accurate than the TE listeners when they heard the Thai stops. However, their discrimination accuracy depended on the type of contrast they heard. An acoustic analysis revealed clearly distinct patterns of formant trajectory movement for two (/a/, /u/) of the three context vowels used in the stimuli. Between the final /p/ and /t/, the F2 offset values for these vowels differed by as much as 330 Hz and 600 Hz for /a/ and /u/, respectively, suggesting that these acoustic differences may be audible to the listeners.
General discussion

The AE and TE speakers demonstrated different patterns of results in perceiving final stops in various stimulus types. The TE participants who lived in English-speaking environments for more than 5 years on average discriminated L2 English stops as accurately as L1 Thai stop contrasts. Their discrimination accuracy of English stops did not differ significantly from that of AE listeners.

Although English final stops are variably released depending on factors such as the vowel type that precedes the final stop (Parker and Walsh, 1981; Lisker, 1999), gender of the talker (Byrd, 1992, 1993, 1994), place of articulation (Crystal and House, 1988; Byrd, 1993), dialect (Byrd, 1992), speaking style (Picheny, Durlach and Braid, 1985, 1986; Bond and Moore, 1994) and the position of the stop within the utterance (Halle, Hughes and Radley, 1957), the occurrence of unreleased stops in an L1 did not appear to assist the AE listeners in perceiving Thai final stops. In order to determine whether their difficulty was due to cross-linguistic acoustic differences between English and Thai or if some other explanations are possible, it is necessary to compare naturally unreleased English stops and Thai stops.

It would also be necessary to examine the perception of English stops by monolingual Thai speakers who do not have much experience in listening to English in order to assess if the TE participants’ perception of English stops can be attributed to their L2 speech learning or to acoustic characteristics of English stops accompanied with a release burst. Perception data recently obtained from native Thai high school students in Bangkok (Tsukada and Roengpitya, under review) suggest that the TE listeners’ discrimination accuracy is a result of their L2 learning.

Finally, it would be interesting to examine the effect of perceptual training on the ultimate accuracy with which non-native sound contrasts are perceived. Although the AE listeners’ perception of Thai stops was dissimilar to that of TE listeners’, they may learn to shift their attention to critical cues and accurately discriminate Thai final stops with much less training than it would take native Japanese listeners to discriminate /l/ and /r/. In other words, the AE listeners’ initial difficulty with Thai stops may not persist.

In sum, native and non-native listeners demonstrated distinct response patterns to Thai and English stop contrasts according to their L1. Although a cross-linguistic comparison in listeners’ L1 phonology (i.e. English and Thai share the same phonemic contrasts in stop sounds) led to the expectation that the AE and TE listeners would discriminate stop contrasts in both English and Thai accurately, clear between-group differences emerged for the Thai, but not English, stimuli. An acoustic examination of the Thai stimuli suggested that the listeners’ higher discrimination accuracy for the /p/-/t/ contrast than for the other contrasts was related to distinct formant trajectory patterns for the vowels /a/ and /u/ when followed by /p/ and /t/. Together, these findings suggest that phonetic realizations of stop sounds and the amount of acoustic information contained in the speech signal may influence accuracy with which sound contrasts are discriminated.

References


Received November 12, 2004
Revision received May 22, 2005; August 30, 2005
Accepted September 30, 2005