CROSS-LANGUAGE PERCEPTION OF WORD-FINAL STOPS: COMPARISON OF CANTONESE, JAPANESE, KOREAN AND VIETNAMESE LISTENERS

Kimiko Tsukada[1] Thu T. A. Nguyen[2] Rungpat Roengpitya[3] Shunichi Ishihara[4]

[1]University of Oregon [2]The University of Queensland [3]Mahidol University
[4]The Australian National University
[1]kimiko.tsukada@gmail.com [2]thunguyen@uq.edu.au [3]rungpat@gmail.com
[4]shunichi.ishihara@anu.edu.au

ABSTRACT

This study examined the discrimination of word-final stop contrasts (/p/-/t/, /p/-/k/, /t/-/k/) in English and Thai by native speakers of Cantonese (C), Japanese (J), Korean (K) and Vietnamese (V). The listeners' first languages (L1) differ substantially in how word-final stops are phonetically realized.

Although Japanese does not permit word-final stops, the J listeners were able to discriminate English (but not Thai) contrasts accurately, demonstrating that non-native contrasts are learnable beyond early childhood. The C, K and V listeners have experience with unreleased final stops in their L1s, but differed in their discrimination accuracy especially for Thai stop contrasts.

This research highlights the value of systematically comparing listener groups from diverse L1 backgrounds in gaining a better understanding of the role of L1 experience in cross-language speech perception.

Keywords: cross-language perception, final stop, Asian languages

1. INTRODUCTION

This study built on recent research [12 13] and compared discrimination accuracy of word-final stops by listeners from four different Asian language backgrounds (Cantonese, Japanese, Korean, Vietnamese) with a view to exploring the role of first language (L1) on cross-language speech perception.

Word-final stops are *sometimes* unreleased in English but *always* unreleased in Thai. Previous research has shown that unreleased English stops are less intelligible than their released counterparts [6 10 14]. However, [1] showed that native Thai speakers were able to accurately identify the stop place of articulation in Thai. On the other hand, native speakers of American English who have no experience with Thai were less accurate than Thai speakers in identifying the Thai stop place [1].

In order to determine if similarity between L1 and the stimulus languages at phonetic as well as phonological levels might be beneficial in cross-language speech perception, adult listeners of Cantonese, Japanese, Korean and Vietnamese were recruited in the present study. Of these four languages, Cantonese, Korean and Vietnamese share specific phonetic realization with Thai, i.e., unreleased final stop ([7-9]). Japanese, on the other hand, does not permit word-final stops.

2. METHOD

In this experiment, cross-language perception by listeners from four Asian language backgrounds was examined. Results from the two control groups (Australian English, Thai) were reported elsewhere [12]. The experimental stimuli and procedures were identical to those in earlier studies.

2.1. Stimuli

The Australian English and Thai speakers read monosyllabic (CVC) words ending in /p t k/ in their L1s in the MARCS Auditory Laboratories recording studio at the University of Western Sydney, Australia. Test words (all real words in English or Thai) were presented to each speaker in randomized orders. Thai words were transcribed using Thai scripts and had either high or low tones. The recorded speech materials were digitized at 44.1 kHz using CoolEdit and amplitude of each sound file was normalized to 50% of the peak following the procedures used in previous research [2 4 5].

Tokens from three female speakers were used for English stimuli and tokens from three male speakers were used for Thai stimuli. More than 90% of the 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[1].

2.2. Listeners

A total of 55 adult listeners participated in this study. They include 13 Cantonese, 12 Korean, 18 Japanese and 12 Vietnamese listeners. These listeners had lived in English-speaking countries (either Australia or USA) for a varying number of years at the time of testing. The participants responded to a local advertisement and received payment for their participation.

2.3. Task

A categorial discrimination test (CDT) employed in previous L2 speech research (e.g., [2 4 5 12 13]) was used. The stimuli (monosyllabic CVC words) were presented in triads via headphones at a self-selected comfortable level using a notebook computer. 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 an odd item out, 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 ‘sip$_2$’-‘sit$_1$’-‘sip$_3$’ (where the subscripts indicate different talkers). The correct response for change trials was the button (“1”, “2”, or “3”) indicating the position of the odd item out. 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/$_3$ /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).

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 English first and then Thai. This was to ensure that they understood the task.

Responses to the change and no-change trials were used to calculate A’ scores[2] [11], an index of discrimination accuracy. A score of 1.0 indicated perfect sensitivity, whereas a score of 0.5 or lower indicated a lack of phonetic sensitivity.

3. RESULTS

Figures 1-4 show mean discrimination scores for English and Thai stimuli by the four groups of listeners as a function of contrast types. All groups were more accurate in discriminating English than Thai contrasts. As described below, listeners’ discrimination accuracy depended on the contrast type, in particular, for the Thai stimuli.

A Group (C, J, K, V) x Language (English, Thai) x Contrast (/p/-/t/, /p/-/k/, /t/-/k/) ANOVA yielded significant effects for main factors [G: $F(3, 51) = 4.5$, $p < 0.01$, L: $F(1, 51) = 117.7$, $p < 0.001$, C: $F(2, 102) = 68.6$, $p < 0.001$] and all three two-way interactions [G x L: $F(3, 51) = 6.4$, $p < 0.001$, C x L: $F(2, 102) = 43.2$, $p < 0.001$, G x C: $F(6, 102) = 2.7$, $p < 0.05$]. A three-way interaction did not reach significance [$F(6, 102) = 1.2$, *ns*].

Figure 1: Mean discrimination scores for Thai and English stimuli by 13 C listeners. The brackets enclose ± one SE.

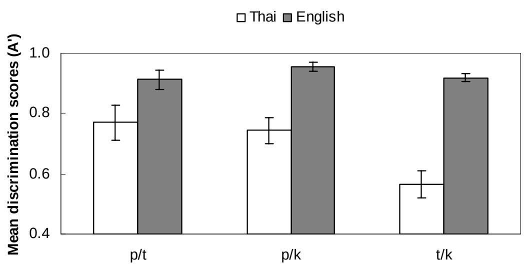


Figure 2: Mean discrimination scores for Thai and English stimuli by 18 J listeners. The brackets enclose ± one SE.

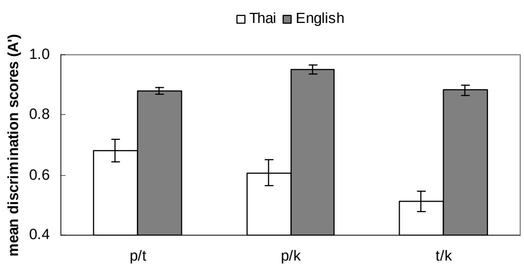


Figure 3: Mean discrimination scores for Thai and English stimuli by 12 K listeners. The brackets enclose ± one SE.

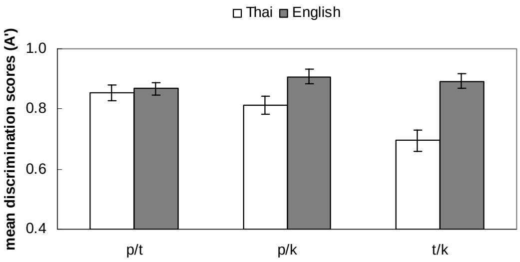
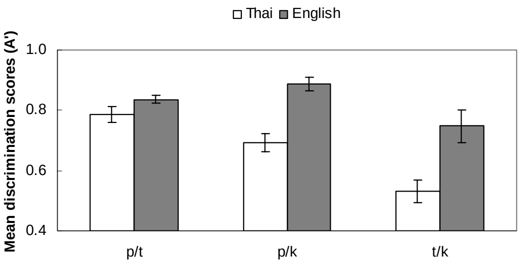


Figure 4: Mean discrimination scores for Thai and English stimuli by 12 V listeners. The brackets enclose ± one SE.

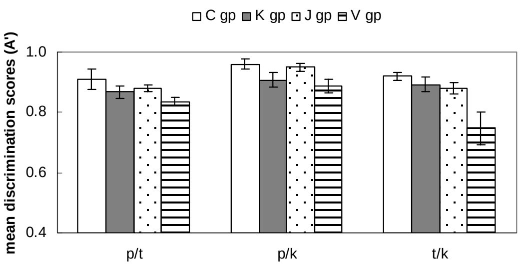


3.1. English stimuli

Figures 5 shows mean discrimination scores for English stimuli by four groups of listeners as a function of contrast types.

A Group x Contrast ANOVA yielded significant effects for both main factors [G: $F(3, 51) = 5.9$, $p < 0.01$, C: $F(2, 102) = 19.3$, $p < 0.001$]. A two-way interaction also reached significance [$F(6, 102) = 3.6$, $p < 0.01$].

Figure 5: Mean discrimination scores for English stimuli by 4 groups of listeners. The brackets enclose ± one SE.

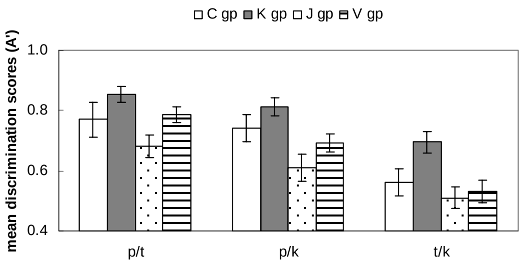


The simple effect of Contrast was significant for C, J and V groups, but not for K group. All four groups discriminated the /p/-/k/ contrast most accurately. The simple effect of Group was significant for /p/-/t/ and /t/-/k/, but not /p/-/k/. For /p/-/t/, the C listeners were better than the V listeners and for /t/-/k/, the C, J and K listeners were better than the V listeners.

3.2. Thai stimuli

Figure 6 shows mean discrimination scores for Thai stimuli by four groups of listeners as a function of contrast types. For the Thai stimuli, there was a clearer between-group difference in the discrimination accuracy compared to the English stimuli in Figure 5.

Figure 6: Mean discrimination scores for Thai stimuli by 4 groups of listeners. The brackets enclose ± one SE.



The main effects of Group and Contrast reached significance [G: $F(3, 51) = 5.1$, $p < 0.01$, C: $F(2, 102) = 72.5$, $p < 0.001$], but a two-way interaction did not [$F(6, 102) = 1.3$, *ns*].

The simple effect of Contrast was significant for all four groups. All groups discriminated the /p/-/t/ contrast most accurately, but they differed slightly in their discrimination accuracy patterns. The simple effect of Group was significant for all contrasts. The K listeners were better than the J listeners for all contrasts and for /t/-/k/, they were better than the V listeners, as well.

4. DISCUSSION

The present study demonstrated that adult native listeners of four Asian languages discriminated final stop contrasts more accurately in English, a language they were familiar with than in Thai, a language unknown to them. While this finding may not be surprising, their results are revealing when their L1 backgrounds are taken into account.

Despite the absence of word-final stops in their L1, the J listeners were able to discriminate English (but not Thai) contrasts accurately. Although neither English nor Thai contrasts are considered phonemically novel to the C, K and V listeners, the extent of L1 benefit differed among those listeners from 'non-release' language backgrounds and hearing allophonic variation of unreleased stops in their L1 did not automatically give them an advantage in discriminating unreleased Thai stops. These findings highlight the value of comparing listener groups from diverse L1 backgrounds.

In fact, the extent to which unreleased final stops are identified or discriminated appears to vary greatly even among native listeners of those 'non-release' languages [1 3 8 9]. If that is the case, L1 listeners' attunement to such acoustic cues differing in robustness may influence their cross-language perception. This needs to be confirmed by acoustic analyses of their L1 stops.

The present findings suggest that specific phonetic experience is not sufficient for accurate perceptual discrimination of certain speech contrasts in an unfamiliar language. Rather, our results seem to suggest that exposure to native input and familiarity with speech materials at both phonological and phonetic levels appears to be needed for accurate discrimination of sounds in an unknown language.

5. CONCLUSIONS

Our tentative conclusion is that experience with specific but non-native phonetic realization of final stops may not be sufficient to ensure accurate discrimination in an unfamiliar language even though the language of interest does not violate listeners' L1 phonetic/phonological specifications. Rather, an early exposure to specific *and* native phonetic contrasts that include detailed acoustic characteristics may be crucial in developing the capacity to discriminate subtle phonetic contrasts as truly native listeners. This highlights the importance of not only the amount but also the *kind* of input in the acquisition of speech sounds.

6. ACKNOWLEDGEMENTS

This research was supported by the University of Western Sydney seed grant and small professional grant from the Center for Asian and Pacific Studies, University of Oregon. Portion of this work was presented at the 153rd ASA Meeting in Salt Lake City in June 2007. We thank the following individuals and organizations: P. Bryan, R. Fahey, C. Kieffer, J. Liang and S. Spinks for subject recruitment and data collection, S. Rungrojsuwan and C. Schoknecht for advice on Thai, the Speech, Hearing and Language Research Centre and Department of Linguistics, Macquarie University for the use of the perception laboratory. We also thank two reviewers for their time.

7. REFERENCES

[1] Abramson, A., Tingsabadh, K. 1999. Thai final stops: Cross-language perception. *Phonetica* 56, 111-122.

[2] Aoyama, K, Flege, J. E., Guion, S. G., Akahane-Yamada, R., Yamada, T. 2004. Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *J. Phon.* 32, 233-250.

[3] Chen M., Wang W. S-Y. 1975. Sound change: actuation and implementation. *Language* 51, 255-281.

[4] Flege, J. E., MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26, 1-34.

[5] Guion, S. G., Flege, J. E., Akahane-Yamada, R., Pruitt, J. C. 2000. An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *J. Acoust. Soc. Am.* 107, 2711-2724.

[6] Householder, F. W. Jr. 1956. Unreleased PTK in American English. In: Halle, M., Lunt, H. G., McLean, H., van Schooneveld, C. (eds.), *For Roman Jacobson*. The Hague: Mouton, 235-244.

[7] Ingram, J. C. L, Pittam, J. 1987. Auditory and acoustic correlates of perceived accent change: Vietnamese schoolchildren acquiring Australian English. *J Phon.* 15, 127-143.

[8] Khouw E., Ciocca V. 2006. An acoustic and perceptual study of final stops produced by profoundly hearing-impaired adolescents. *J. Sp. Lang. Hear Res.* 49, 172-185.

[9] Law S-P, Fung R. S-Y, Bauer R. 2001. Perception and production of Cantonese consonant endings. *Asia Pacific J. Sp. Lang. Hear.* 6, 179-195.

[10] Malécot, A. 1958. The role of releases in the identification of released final stops. *Language* 34, 370-380.

[11] Snodgrass, J. G., Levy-Berger, G., Haydon, M. 1985. *Human Experimental Psychology.* New York: Oxford University Press.

[12] Tsukada, K. 2006. Cross-language perception of final stops in Thai and English. *Bilingualism: Language and Cognition* 9, 309-318.

[13] Tsukada, K., Ishihara, S. 2007. The effect of first language (L1) in cross-language speech perception: Comparison of word-final stop discrimination by English, Japanese and Thai listeners. *J. Phon. Soc. Japan,* 11, 82-92.

[14] Wang, W. S-Y. (1959). Transition and release as perceptual cues for final plosives. *J. of Sp. Lang. and Hearing Res.* 2, 66-73.

[1] The presence or absence of release burst was determined using a combination of auditory and acoustic information (i.e., inspection of wide-band spectrograms and time domain waveforms).

[2] These scores were based on the proportion of 'hits (H)' obtained for each contrast and the proportion of 'false alarms (FA)'. If H equaled the proportion of 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)). If FA exceeded H, then A' = 0.5-((FA-H)*(1+FA-H))/((4*FA)*(1-H)).