Cross-language speech perception of final stops by Australian-English, Japanese and Thai listeners

Kimiko Tsukada

MARCS Auditory Laboratories
University of Western Sydney, Australia
k.tsukada@uws.edu.au

Abstract

This study examined the discrimination of word-final stop contrasts (/p/-/t/, /p/-/k/, /t/-/k/) in English and Thai by three groups of listeners differing in their first language (L1): Australian English (AE), Japanese (NJ) and Thai (NT). Thai final stops are *invariably unreleased* whereas English final stops are *variably released*. Although Japanese listeners had no experience with word-final stops in the L1, they were able to discriminate English (but not Thai) contrasts accurately, demonstrating that non-native contrasts are learnable and that some aspects of speech perception remain plastic beyond early childhood. The AE listeners did not match the NT listeners in discriminating Thai stops despite their experience with unreleased stops in the L1. This suggests that phonetic realization of sounds and/or the amount of acoustic information in the speech signal may influence accuracy with which non-native contrasts are discriminated.

1. Introduction

It is well known that the native or first language (L1) phonology exerts a strong influence on the production/perception of non-native sounds. Theories of cross-language speech perception, e.g., [1, 2], posit that the perceived relation between the non-native sounds and the closest native counterparts plays an important role in predicting discrimination accuracy of non-native sound contrasts. According to the Perceptual Assimilation Model [1], instances of contrastive non-native categories that are identified as instances of a single L1 category (single-category assimilation) will be relatively difficult to discriminate whereas instances of contrastive non-native categories that are mapped onto two different L1 categories (two-category assimilation) will be discriminated more accurately.

There is evidence that L1 phonology alone does not completely predict listeners' response patterns for non-native sound contrasts. While Japanese L2 learners are known to produce and perceive English /l/ and /r/ poorly, their discrimination accuracy depended on the cross-linguistic phonemic assimilation patterns and listeners' experience with spoken English [3]. Although Japanese does not permit consonants in syllable-final position (except for a few cases), the L2 learners tested were able to distinguish /l/ and /r/ more accurately in final than initial position.

A recent study examined the perception of syllable-initial and syllable-final English nasals by Korean and Japanese L2 learners [4]. The NJ 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.

Conversely, L1 phonology may overestimate listeners' responses to non-native sound categories [5]. For instance, four English approximants /w j r l/ occur in French with varying degrees of similarity to the target sounds. If equivalence at the traditional phonological level predicts 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.

This study examined the ability of three groups of listeners differing in L1 to discriminate native and/or non-native stops in word-final position. Native speakers of Australian English (AE) and Japanese (NJ) and Thai (NT) speakers living in Australia with varying degrees of experience with English participated and their discrimination of three stop contrasts (/p/-/t/, /p/-/k/, /t/-/k/) in English and Thai was assessed.

The listeners' L1s differed markedly with respect to the occurrence of word-final stops. All stop contrasts tested here occur in English and Thai and are known/familiar to these listeners but they differ considerably in phonetic realization, i.e., *variably released* in English and *invariably unreleased* in Thai. Previous research has shown that unreleased stops are less intelligible than released stops [6, 7, 8]. However, NT speakers were able to accurately identify the stop place of articulation in Thai and native speakers of American English were less accurate than NT speakers in identifying the Thai stop place [9].

In Japanese, the predominant syllable structure is /(C)V/ and English words such as 'cap' and 'cat' are expected to be difficult for Japanese listeners to discriminate. There are numerous studies that examined the acquisition of English sounds by NJ learners and it is very well known that Japanese L2 learners have considerable difficulty producing and perceiving English /l/ and /r/ (see [10] for a review). However, very little published research seems available for their acquisition of final stops contrasting in place of articulation (see, however, [4] for the perception of English nasal contrasts). This is even more surprising given the basic syllable structure in Japanese (i.e., /(C)V/) that does not permit syllable-final consonants and potential difficulties for

NJ learners of English to produce and/or perceive contrasts in final position.

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. Two-month old infants were shown to be sensitive to place of articulation differences in CV<u>C</u> and V<u>C</u> syllable pairs (target sounds underlined) and that they did not have any more difficulty with stop contrasts in final position than in initial position [11]. 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 2 month, infants in an English-speaking environment were learning to discriminate contrasts in both syllable positions that are legitimate in their L1.

The cross-linguistic differences reviewed above lead to the expectations that AE and NT listeners would discriminate stop contrasts in English and Thai accurately while NJ listeners would discriminate both of them poorly. Specific questions addressed in this study are as follows: 1) Do Australian English and Thai listeners whose L1s permit word-final stops show accurate discrimination of stop contrasts in non-native languages? 2) Do native Japanese listeners whose L1 does not permit word-final stops show poor discrimination of stop contrasts in English and Thai?

2. Method

2.1. Stimuli

Stop pairs (/p/-/t/, /p/-/k/, /t/-/k/) contrasting in places of articulation in both English and Thai were used as stimuli. Participants read monosyllabic (/CVC/) words ending in /p t k/ in their L1s in the MARCS recording studio at 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, amplitude normalized to 50% of the peak and stored in separate files per word. In order to minimize the speaker gender 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. Nearly all 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.

2.2. Listeners

Three groups of listeners participated: native speakers of Australian English (n=10), Thai (n=26) and Japanese (n=4)[1]. All participants reported normal hearing and had no history of language problems. Mean characteristics of the three groups of participants are summarized in Table 1.

Table 1: Mean characteristics of three groups of participants

Group	n	CA	AOA	LOR
AE	1m, 9f	23.4 (9.3)	--	--
NT	6m, 20f	29.8 (10.6)	24.6 (6.6)	5.3 (6.0)
NJ	4f	30.4 (5.6)	27.8 (4.2)	2.8 (2.2)

CA = chronological age in years
AOA = age of arrival in Australia in years
LOR = length of residence in Australia in years
Standard deviations are in parentheses.

Ten English-speaking listeners were enrolled in first-year Psychology course at University of Western Sydney. The AE and NT participants (except for 3 NT listeners who were tested in Canberra) were tested in MARCS Auditory Laboratories. The NJ participants were enrolled in a MA in Translation and Interpreting at Macquarie University. None of the AE and NJ listeners had any knowledge of Thai. All participants were tested individually in a single session lasting about 45 minutes.

2.3. Task

A categorial discrimination test (CDT) employed in previous L2 speech research, e.g., [12, 13, 14, 15], was used. Monosyllabic (/CVC/) words ending with stop tokens were presented in triads to listeners whose task was to identify an odd item out if any.

The stimuli were presented 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 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 'sip_2'-'sit_1'-'sip_3' (where the subscripts indicate different talkers). In other words, in change trials, listeners would hear minimal pairs differing only in the place of articulation of the 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 $/t/_3$ $/t/_1$ $/t/_2$), 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). The participants were required to respond to each trial, and were told to guess if uncertain. A trial could be

[1] Additional AE and Japanese listeners were being recruited and tested at the time of submitting this paper so that the number of listeners in each group would be more balanced. Their results are currently being analyzed.

replayed, but responses could not be changed once given. The inter-stimulus interval in all trials was 0.5 s. The first ten tokens were for practice and were not analyzed. The AE and NT listeners heard stop contrasts in their own language first and then in the other language. The NJ listeners heard English stimuli first and then Thai stimuli.

In order to reduce possible effects of response bias, the dependent variable examined in all analyses were A' scores based on responses to both change and no-change trials [16]. 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)). However, if FA exceeded H, then A' = 0.5-((FA-H)*(1+FA-H))/((4*FA)*(1-H)). A score of 1.0 indicated perfect sensitivity, whereas a score of 0.5 or lower indicated a lack of phonetic sensitivity.

3. Results

All three groups of listeners showed accurate discrimination of the English stop contrasts, but only the NT listeners were able to discriminate the Thai stop contrasts accurately. The AE listeners discriminated the Thai stops more accurately than the NJ listeners but less accurately than the NT listeners. The listeners' discrimination accuracy differed somewhat according to the contrast type. Two-way ANOVAs with Group (AE, NJ, NT) as a between-subjects factor and Contrast (/p/-/t/, /p/-/k/, /t/-/k/) as a within-subjects factor were carried out separately for each stimulus language.

3.1. English stimuli

Figure 1 shows the results when three groups of listeners heard the English stimuli. All groups obtained the mean scores higher than 0.85 for all contrasts tested. The effect of Group did not reach significance, but the effect of Contrast did [$F(2, 74) = 9.2$, $p < 0.001$]. All three groups discriminated the /p/-/k/ contrast more accurately than the other two contrasts. It has been shown that the intelligibility of the final /k/ is particularly susceptible to the presence or absence of a release burst [9, 17, 18]. 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 listener groups responded in a comparable manner to different stop contrasts.

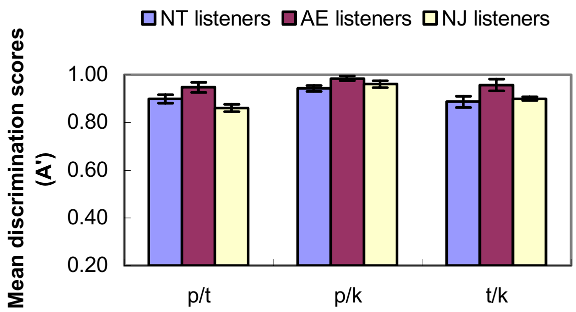

Figure 1: Mean discrimination scores for the English stimuli by three groups of listeners. The brackets enclose ± one standard error.

3.2. Thai stimuli

Figure 2 shows the results when three groups of listeners heard the Thai stimuli.

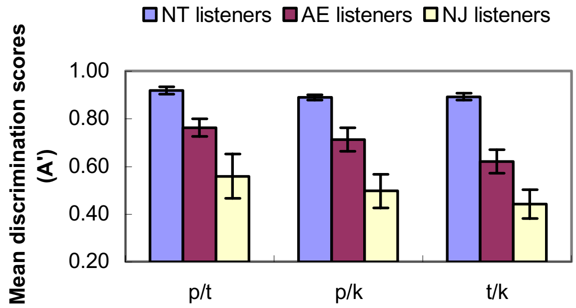

Figure 2: Mean discrimination scores for the Thai stimuli by three groups of listeners. The brackets enclose ± one standard error.

Averaged across three contrasts, the NT group scored 0.90 as opposed to 0.70 and 0.50 for the AE and NJ groups, respectively. The two non-native groups were also more variable than the NT group. The ANOVA yielded significant main effects of Group [$F(2, 37) = 48.5$, $p < 0.001$] and Contrast [$F(2, 74) = 11.5$, $p < 0.001$]. A two-way interaction also reached significance [$F(4, 74) = 3.4$, $p < 0.05$], which was explored by simple effects tests.

The simple effect of Group was significant for all contrasts [/p/-/t/: $F(2, 37) = 25.6$, $p < 0.001$; /p/-/k/: $F(2, 37) = 33.6$, $p < 0.001$; /t/-/k/: $F(2, 37) = 46.6$, $p < 0.001$]. The advantage of the NT group over the other groups can be clearly seen in Figure 2 for all stop contrasts tested. The AE group was significantly better than the NJ group for all contrasts, but worse than the NT group.

The simple effect of Contrast was significant for the AE and NT groups [AE: $F(2, 18) = 3.8$, $p < 0.05$; NT: $F(2, 50) = 4.3$, $p < 0.05$], but not for the NJ group. For the AE group, the discrimination accuracy was highest for /p/-/t/, lowest for /t/-/k/ and intermediate for /p/-/k/. The NT listeners were also most accurate for the /p/-/t/ contrast, but unlike the AE group, they were equally accurate in discriminating the other two contrasts. All three groups discriminated the /p/-/t/ contrast most accurately. Thus, a significant interaction arose due to differing patterns of Contrast and Group effects.

4. Discussion

This study examined the discrimination of word-final stop contrasts by listeners differing in their L1s. It was clearly shown that a correspondence or equivalence at the phonemic level was insufficient for accurate discrimination of stop contrasts. The three groups of listeners did not differ in their discrimination of the English stop contrasts, but they diverged when they heard the Thai stop contrasts.

The NT listeners who are experienced with unreleased final stops in their L1 showed accurate discrimination of stop contrasts in both English and Thai. The NT group showed more accurate discrimination of the Thai stop contrasts than the AE group who, in turn, performed better than did the NJ group. It is to be noted that unreleased stops do not violate the phonological rules of English.

It was interesting that the difference between the AE and NJ groups emerged only in their discrimination accuracy of an unfamiliar language, Thai. Although it appeared that the AE listeners did not fully benefit from the experience of hearing unreleased stops in their L1 when they heard the Thai stops, it may still be due to their experience with unreleased English stops that they were better able than the NJ listeners to extract relevant cues to discriminate the Thai stop contrasts.

The lack of between-group difference for the English stimuli may be because the listeners were at ceiling. All the NJ listeners tested were living in Sydney for a varying period of time (1.4 – 6 years) and, although they were late L2 learners, they were not inexperienced in English. NJ listeners with no or very limited experience with spoken English may respond to the same stimuli in a very different manner. If such results were obtained, it would confirm that the NJ listeners in the present study have considerably modified their perceptual behaviours. It is still possible that released English stops are highly intelligible to any listener regardless of language-specific experience.

In this connection, an investigation is currently underway with NT listeners in Bangkok who show variation in their English experience. So far, preliminary results of 8 listeners, all majoring in Linguistics at a university, were exactly like the results obtained for the NT participants in this study. These listeners discriminated final stops accurately in both English and Thai.

5. Conclusions

Although the NJ listeners had no experience with word-final stops in the L1, they were able to discriminate English (but not Thai) stop contrasts accurately, demonstrating that non-native contrasts are learnable and that some aspects of speech perception remain plastic beyond early childhood. Taken together with the finding that the AE listeners did not match the NT listeners in discriminating Thai stops despite their experience with unreleased stops in the L1, it is suggested that phonetic realization of sounds and/or the amount of acoustic information contained in the speech signal may influence accuracy with which sound contrasts are discriminated.

6. Acknowledgements

This research was supported by the UWS seed grant. My conference attendance was financially supported by ASSTA Travel Award. I thank the SHLRC (Speech, Hearing and Language Research Centre) and Department of Linguistics at Macquarie University for the use of the perception laboratory. I also thank S. Ishihahra and J. Liang for subject recruitment and data collection, and R. Roengpitya, S. Rungrojsuwan and C. Schoknecht for advice on Thai speech materials.

7. References

[1] Best, C. T. (1995). A direct realist view of cross-language speech. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 171-204). Timonium, MD: York Press.

[2] Flege, J. E. (1995) Second language speech learning: theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.

[3] Best, C., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics,* 20, 305-330.

[4] Aoyama, K. (2003). Perception of syllable-initial and syllable-final nasals in English by Korean and Japanese speakers. *Second Language Research,* 19, 251-265.

[5] Hallé, P. A., Best, C. T., & Levitt, A. (1999). Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics,* 27, 281-306.

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

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

[8] Wang, W. S-Y. (1959). Transition and release as perceptual cues for final plosives. *Journal of Speech, Language and Hearing Research,* 2, 66-73.

[9] Abramson, A. S., & Tingsabadh, K. (1999). Thai final stops: cross-language perception. *Phonetica,* 56, 111-122.

[10] Aoyama, K., Flege, J. E., Guion, S., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics,* 32, 233-250.

[11] Jusczyk, P. (1977). Perception of syllable-final stop consonants by 2-month-old infants. *Perception & Psychophysics,* 21, 450-454.

[12] Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America,* 106, 2973-2987.

[13] 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. *Journal of the Acoustical Society of America,* 107, 2711-2724.

[14] Flege, J.E. (2002). Methods for assessing the perception of vowels in a second language. In E. Fava E & A. Mioni (eds.), *Issues in clinical linguistics.* pp. 3-28. Padova: UniPress.

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

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

[17] Lisker, L. (1999). Perceiving final voiceless stops without release: effects of preceding monophthongs versus nonmonophthongs. *Phonetica,* 56, 44-55.

[18] Tsukada, K., Birdsong, D., Mack, M., Sung, H., Bialystok, E., & Flege, J. (2004). Release bursts in English word-final voiceless stops produced by native English and Korean adults and children. *Phonetica,* 61, 67-83.