Factors affecting talker adaptation in a second language

Anne Cutler, L. Ann Burchfield, and Mark Antoniou

The MARCS Institute for Brain, Behaviour and Development, Western Sydney University

a.cutler/a.burchfield/m.antoniou@westernsydney.edu.au

Abstract

Listeners adapt rapidly to previously unheard talkers by adjusting phoneme categories using lexical knowledge, in a process termed lexically-guided perceptual learning. Although this is firmly established for listening in the native language (L1), perceptual flexibility in second languages (L2) is as yet less well understood. We report two experiments examining L1 and L2 perceptual learning, the first in Mandarin-English late bilinguals, the second in Australian learners of Mandarin. Both studies showed stronger learning in L1; in L2, however, learning appeared for the English-L1 group but not for the Mandarin-L1 group. Phonological mapping differences from the L1 to the L2 are suggested as the reason for this result.

Index Terms: speech perception, perceptual learning, Mandarin, English, second language learning

1. Introduction

Human listeners adapt to newly-encountered talkers with remarkable rapidity. In the past decade and a half, this process has been extensively investigated using a paradigm in which listeners hear ambiguous phonetic forms which they are able to disambiguate by reference to existing knowledge (see [1] for a review). The initial use of this paradigm [2] established that exposure to just 20 instances of a deviant phonemic form induces learning about the speaker's putative pronunciation of the sound in question, as long as the deviant form is heard in real-word contexts so that it can be ascribed to a phonemic category. Thus an ambiguous sound between /s/ and /f/ will be learned as /s/ if heard in words like *horse*, as /f/ if heard in words like *giraffe*, but will remain ambiguous if heard in nonwords such as *liff* or *liss*. The learning generalises to other words containing the same phoneme, creating a path for rapid adaptation on first exposure to speech from a new talker.

The perceptual learning process has been documented for numerous types of phoneme, in different positions in the word, and for listeners from childhood to old age. It has been shown to be rapid (less than 20 exemplars also work), in good part speaker-specific though with reasonable generalisations across phoneme class and language variety, and long-lasting; further, it can be induced by exposure to words in context or in isolation, and can be measured in either phonetic decision tasks or lexical disambiguation tasks (on all counts see [1]).

Of particular relevance to the present work is that this rapid learning has been observed in many languages, both European and non-European (and indeed that it also holds for lexically distinctive non-segmental speech sounds; in Mandarin, for example, a speaker's use of a lexical tone ambiguous between Mandarin tones 1 and 2 led to adjustment [3] in the same way as was also seen for ambiguous phonemes in Mandarin [4]). Furthermore, the learning can be successfully applied in a second language [5,6], although this has chiefly been observed in related L1 and L2 and in an immersion environment.

In the present study we addressed the relative ability of listeners to apply this rapid adjustment in their first and second languages, and we chose a pair of languages with a high degree of phonological dissimilarity: Mandarin Chinese and (Australian) English. Perceptual learning has already been demonstrated for these languages [4,7], so we expect each L1 to exhibit a robust effect. However, the L2 cases may differ. One of these is an immersion environment: Experiment 1, in which we examined the learning effect in the L2 of Mandarin-English late bilinguals living in Australia. The other is a case without immersion: Experiment 2, where we examined the same learning for L2 in Australian adult learners of Mandarin, still resident in their English-speaking native environment.

2. Pilot experiment

To select an ambiguous fricative for the main experiments, a pilot experiment was conducted in which participants heard and categorised steps from an [f]-[s] continuum in English and Mandarin. A female native speaker of each of Mandarin and English produced the syllables /fu/, /su/, and /θu/. The fricative portions of the /fu/ and /su/ recordings were excised and a 41-step continuum was created in each language (following [2]). Using Praat [8], [f] and [s] waveforms were mixed in constant proportions along a 41-step continuum such that one end of the continuum was 100% [f], 0% [s] and the other end 0% [f], 100% [s]. Fricatives were spliced onto the vowel /u/ taken from the same speaker's production of /θu/. This avoided coarticulatory cues in vowels biasing listeners to interpret the ambiguous sounds as either [f] or [s]. Fourteen steps were chosen from this [f]-[s] continuum as stimuli for this pilot: 1 ([f]), 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29 and 41 ([s]). For each continuum (Mandarin, English), a separate group of native listeners heard these 14 steps 10 times in random order and categorised each token by pressing "F" or "S" on a computer keyboard.

Mandarin and English results from the pilot are shown separately in Figure 1. In both languages, listeners' responses proved step 17 to be the most ambiguous token of the 14 steps tested, and thus step 17 was in each case used to construct the ambiguous stimuli for the perceptual learning training trials.

3. Experiment 1

3.1. Method

3.1.1. *Participants*

24 Mandarin-English late bilingual speakers (mean age 26.6, range = 21.6-36.3; mean age of arrival in Australia 23.3, range = 15.2-35.4; mean length of residence 3.2 years, range = 0.5-9.6) took part in return for a small payment. All reported Mandarin Chinese as their dominant language, and did not report pre-school-age exposure to other languages or dialects. None reported any vision, hearing, or language impairments.

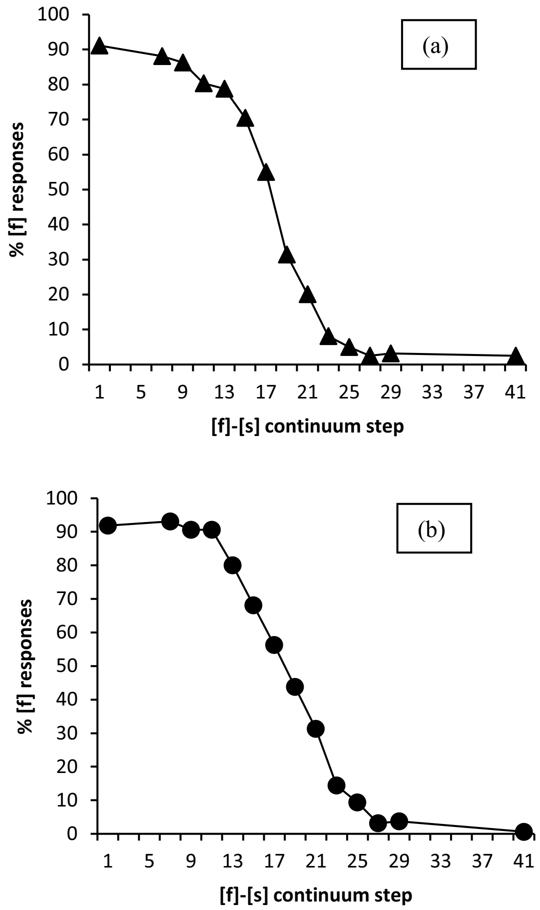

Figure 1: *Pilot experiment: Total proportion of [f] responses for (a) Mandarin and (b) English. In each case, step 17 was chosen as the ambiguous sound for training trials.*

3.1.2. Training stimuli

The training materials for each language were all disyllabic: 100 words and 100 non-words. 60 words were filler items and 40 were training items. Half of the latter were f-words (with [f] as first phoneme of syllable 2, e.g. *bu4fa3* 'illegal'; *traffic*) and half were s-words (with [s] in that position: *kuan1song1* 'loose'; *gossip*). Words were chosen such that using the other fricative would yield a nonword. The mean frequency for Mandarin words was 3.68 and 3.82 per million for f-words and s-words respectively (computed from the online CCL corpus of PKU [9]). The mean frequency for English words was 4.3 and 4.1 respectively for f-words and s-words (computed from SUBTLEX using the Zipf scale [10]).

For each language two versions of each training item were selected: one unaltered, and another with the critical word-medial fricative replaced by an ambiguous sound [?] (step 17 of the /fu/-/su/ continuum from the pre-test). 100 nonwords were created (in Mandarin this was by changing the tone of the second syllable in a real word (e.g., *ji1-dan4* 'egg' became *ji1-dan2*). Nonwords and fillers did not contain [f], [s], or [ʂ], [ɕ], [ts], or [tʃ] (to avoid perceptually similar sounds to the critical fricatives). The training materials were produced by the same speaker as for the [f]-[s] continuum.

3.1.3. Procedure

Participants completed two full sessions (lexically-guided perceptual learning followed by test), one in L1 and one in L2. These sessions were spaced 2-3 weeks apart. For each language, participants were first exposed to an ambiguous sound [?] either in f-words or s-words in a lexical decision task. They had to decide whether each item was a real word or a non-word, indicating their response via a button press, with "Yes" responses made using the dominant hand. Four stimulus lists were constructed, each containing the same 100 words and 100 non-words. Items were presented randomly with the restriction that no more than four words or non-words occurred in sequence. Two versions of each presentation order were created, one in which [?] replaced all instances of [f] ([f]-ambiguous group) and one in which [?] replaced all instances of [s] ([s]-ambiguous group); half of the participants were assigned to each group. The first 12 trials contained no instances of [?] and were identical across all versions and lists.

Following the lexical decision task, participants completed a categorisation task, in which they heard recordings of steps 7, 13, 17, 21, and 27 of the /fu/-/su/ continuum (identical steps were used for the Mandarin and English continua) and had to categorize each item as either /fu/ or /su/. These five steps were each presented randomly 30 times (150 trials in total).

For each categorisation task, perceptual learning was examined via a 2 × 5 ANOVA with the between-subjects factor of training group ([f]-ambiguous versus [s]-ambiguous) and the within-subjects factor of step (7, 13, 17, 21, 27).

3.2. Results and discussion

3.2.1. Mandarin categorisation

The analysis of the Mandarin-English late bilinguals' perceptual learning of Mandarin, their L1, showed as expected a significant main effect of group, $F(1, 20) = 6.2$, $p = .022$, $\eta_p^2 = .235$ (Figure 2). There was a significant main effect of step, $F(1.9, 38.7) = 199.0$, $p < .001$, $\eta_p^2 = .909$. There was a significant Group × Step interaction, $F(1.9, 38.7) = 6.1$, $p = .005$, $\eta_p^2 = .234$. We examined the interaction via a series of *t*-tests. Training groups differed on steps 13, $t(15.5) = 2.7$, $p = .016$, and 17, $t(20) = 3.5$, $p = .002$.

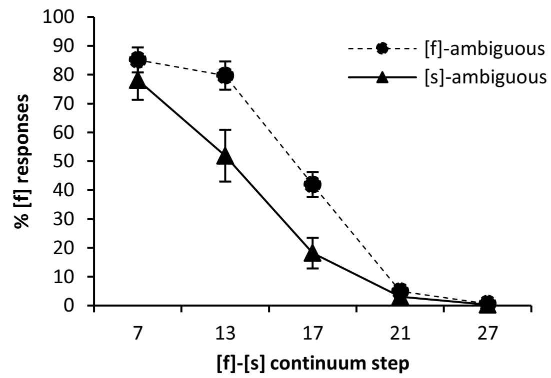

Figure 2: *Total proportion of [f] responses to a Mandarin [fu]-[su] continuum made by Mandarin-English late bilinguals following [f]-ambiguous or [s]-ambiguous training.*

3.2.2. English categorisation

The Mandarin-English late bilinguals' perceptual learning in English showed, in contrast to the L1 results, no significant main effect of group, $F(1, 20) = 0.4$, $p = .541$, $\eta_p^2 = .019$ (see Figure 3). There was a significant main effect of step, $F(2.5, 49.9) = 91.0$, $p < .001$, $\eta_p^2 = .820$, but no significant Group × Step interaction, $F(2.5, 49.9) = 0.6$, $p = .671$, $\eta_p^2 = .029$.

Thus for these listeners, their L1 performance replicated that found in previous work with Mandarin, but in their L2 they showed no evidence of successful perceptual learning.

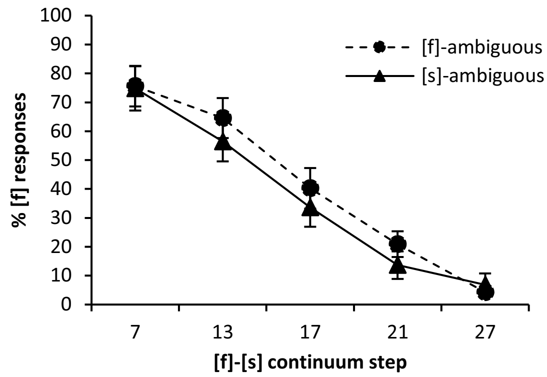

Figure 3: *Total proportion of [f] responses to an English [fu]-[su] continuum made by Mandarin-English late bilinguals following [f]-ambiguous or [s]-ambiguous training.*

4. Experiment 2

4.1. Method

4.1.1. Participants

Experiment 2 involved 25 Australian learners of Mandarin, again paid for participating. These participants were all born in Australia, had a mean age of 30.1 (range = 19.0-54.8), and had acquired Mandarin from a mean age of 14.9 (range = 6-29). All had Australian English as their dominant language. None reported any vision, hearing, or language impairments.

4.1.2. Stimuli and Procedure

These were as in Experiment 1.

4.2. Results and discussion

4.2.1. English categorisation

The Australian Mandarin learners' perceptual learning in their L1 English again showed the expected significant main effect of group, $F(1, 23) = 8.7$, $p = .007$, $\eta_p^2 = .274$ (see Figure 4). There was also a significant main effect of step, $F(2.7, 62.5) = 103.1$, $p < .001$, $\eta_p^2 = .818$, but there was no significant Group × Step interaction, $F(2.7, 62.5) = 1.5$, $p = .222$, $\eta_p^2 = .062$. We examined the group difference via a series of *t*-tests. Training groups differed on step 13, $t(23) = 2.2$, $p = .038$, step 17, $t(23) = 2.5$, $p = .018$, and step 21, $t(13.4) = 3.0$, $p = .010$. This result thus again replicates the perceptual learning effect in English, and replicates the significant learning observed so far in all experiments in an L1 in the L1 environment.

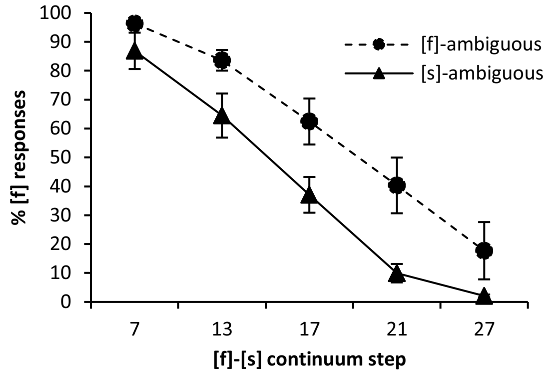

Figure 4: *Total proportion of [f] responses to an English [fu]-[su] continuum made by Australian learners of Mandarin following [f]-ambiguous or [s]-ambiguous training.*

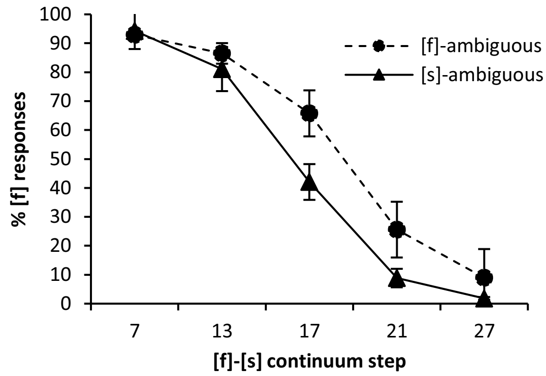

Figure 5: *Total proportion of [f] responses to a Mandarin [fu]-[su] continuum made by Australian learners of Mandarin following [f]-ambiguous or [s]-ambiguous training.*

4.2.2. Mandarin categorisation

The Australian Mandarin learners' perceptual learning in their L2 Mandarin showed a main effect of group that was marginally significant, $F(1, 23) = 3.3$, $p = .081$, $\eta_p^2 = .127$. There was a significant main effect of step, $F(2.9, 66.2) = 178.8$, $p < .001$, $\eta_p^2 = .886$. There was a significant Group × Step interaction, $F(2.9, 66.2) = 2.9$, $p = .046$, $\eta_p^2 = .110$. We examined the interaction via a series of *t*-tests. Training groups differed on step 17, $t(23) = 2.6$, $p = .017$, and the difference for step 21 was marginally significant, $t(12.8) = 1.824$, $p = .092$. Figure 5 shows these results.

Here again the results were weaker for these listeners' L2 than for their L1. However, unlike the Experiment 1 group, this learner population did show evidence of successful adaptation in their L2; the ambiguous sound that they had been trained on was categorised differently depending on their training (though they were not able to generalise the learning as widely across the phonemic continuum as they had done in their L1). The two learner groups that we have tested thus produced differing result patterns, with this difference being furthermore in the opposite direction from what previous L2 perceptual learning findings in the literature might have predicted: an immersion group was here less successful.

5. General Discussion

Our two experiments have revealed an asymmetry in the degree of perceptual learning achieved by L2 listeners within the language pair English-Mandarin Chinese. As expected from extensive prior investigations in other languages, both sets of listeners were able to adjust the [f]/[s] category boundary within their native language to adapt to an apparent talker idiosyncrasy in one of those phonemes. However in their L2 they were less successful, with the Mandarin listeners to English as their L2 performing worse than the Australian English listeners to Mandarin as their L2.

The dissimilarity between these two languages has allowed us to rule out speculations that perceptual learning would only be observed in L2 in the case of related L1/ L2 pairs. English and Mandarin are from different, quite unrelated, language families, and also differ on many independent phonological dimensions of relevance to the listening task. Mandarin has fewer vowels and fewer consonants than English, and it has simpler syllable structure and uses no morphological affixes. In all these respects its phonology is less complex than that of English. In the suprasegmental domain, however, Mandarin also has lexical tones that distinguish words, in which respect its phonology may be held to be more complex than English.

Usefully, with this pair we were able to exploit the same phonemic comparison ([f] versus [s]) as tested in many previous demonstrations of perceptual learning, from [2] on. Fricative perception typically leads to less markedly categorical functions than are seen with other consonants [11], but this is stable across languages. Fricative perception does differ across languages due to fricative inventory size and composition [12], but both Mandarin and English have, by world standards, relatively large fricative repertoires, albeit with somewhat greater competition for [f] in English and for [s] in Mandarin. Our results suggest that language similarity is not a prerequisite for the appearance of L2 perceptual learning.

Also, our study has provided new evidence on the role of linguistic immersion. Previous research showing perceptual learning in L2 had mostly been carried out in situations where the listeners were currently living in an environment in which their L2 was the expected language. Thus German students in the Netherlands showed a perceptual learning effect with Dutch input that was equivalent to that shown by L1 Dutch-speakers hearing the same training materials [5]. Dutch-born emigres in Australia showed perceptual learning in their L2 English (which for many of them had become their dominant language) to a more significant degree than in their original L1, Dutch [6]. Not only were both of these results drawing on a situation of immersion in the L2 linguistic environment, but they also involved related languages which are phonologically quite similar: Dutch, German, English. With the same similar languages, perceptual learning has also been reported with no immersion; [13] for German-English, [14] for Dutch-English.

However, given the fact that the listeners in our study who showed less evidence of perceptual learning (Experiment 1) were those in an immersion environment, while the more successful learners were those without immersion (Experiment 2), we no longer regard immersion as the most crucial factor determining the outcome of such studies. Instead, we suggest that future research should concentrate on determining the exact role of phonological asymmetries in the L1:L2 comparison. Although our stimuli were matched in both phonologies (with the critical phoneme in a word-medial syllable onset in each case), the two languages were, as described above, equally carefully chosen for their lack of phonological match.

Two testable hypotheses seem to merit attention regarding transfer to a second language of one's native facility with adaptation to new talkers via perceptual learning (recall the widespread success in this task across L1 investigations). Either of them could be responsible for the pattern of results we have found. First, it may be the case that transfer of this skill to a new phonological system is <u>easier</u> if the new system is simpler than that of the L1 (as in many ways the Mandarin system is simpler than the English system, given that English has more complex syllables, with both onset and coda clusters, plus morphological affixes on words, etc.). Alternatively, it may be the case that transfer is <u>harder</u> if the new system lacks certain elements that are crucial features of the L1 (as is the case, for instance, with English's lack of tones, which in Mandarin have been shown to be a suitable ground for perceptual learning [3]), or if the new system just has a larger phonemic inventory than the L1 (as English has, in comparison to Mandarin). Each of these hypotheses is testable using further language pairs.

6. Acknowledgements

This study was funded by the Australian Research Council (DP140104389, with additional support from the ARC Centre of Excellence in the Dynamics of Language; CE140100041).

7. References

[1] Cutler, A, Eisner, F, McQueen, J. M. and Norris, D. "How abstract phonemic categories are necessary for coping with speaker-related variation", in Papers in Laboratory Phonology, vol. 10, C. Fougeron, B. Kühnert, M.P. d'Imperio and N. Vallée, Eds. Berlin: Mouton de Gruyter, 2010, pp. 91-111.

[2] Norris, D. McQueen J.M. and Cutler A. "Perceptual learning in speech", Cognit. Psychol., 47: 204-238, 2003.

[3] Mitterer H. Chen, Y. and Zhou X., "Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm", Cognit. Sci., 35: 184-197, 2010.

[4] Burchfield, L.A., Luk, S-H. K., Antoniou, M. and Cutler, A. "Lexically guided perceptual learning in Mandarin Chinese", Proc. Interspeech 2017, Stockholm, pp. 575-580, 2017.

[5] Reinisch, E., Weber, A. and Mitterer, H. "Listeners retune phoneme categories across languages". J. Exp. Psychol: Hum Percept. Perform., 39: 75-86, 2013.

[6] Bruggeman, L. "Nativeness, dominance, and the flexibility of listening to spoken language." PhD thesis, Western Sydney University, 2016.

[7] McQueen, J. M. Tyler, M. D. and Cutler, A. "Lexical retuning of children's speech perception: Evidence for knowledge about words' component sounds." Lang. Learn. Dev., 8: 317-339, 2012.

[8] Boersma, P. and Weenink, D. "Praat: Doing phonetics by computer" (Version 6.0.27) [Computer program; Available: http://www.praat.org/], 2017.

[9] Center for Chinese Linguistics PKU. 语料库检索系统 (网络版) http://ccl.pku.edu.cn:8080/ccl_corpus/ [Online]. Accessed March 19, 2017.

[10] Van Heuven, W. J. B., Mandera, P., Keuleers, E. and Brysbaert, M. "Subtlex-UK: A new and improved word frequency database for British English". Quart. J. Exp. Psychol., 67:1176-1190, 2014.

[11] Repp, B.H. "Two strategies in fricative discrimination", Perc. Psychophys., 30: 217-227, 1981.

[12] Wagner, A., Ernestus, M. and Cutler, A. "Formant transitions in fricative identification: The role of native fricative inventory". J. Acoust. Soc. Am., 120: 2267-2277. 2006.

[13] Schuhmann, K.S. "Perceptual learning in second language learners." PhD thesis, Stony Brook University, 2014.

[14] Drozdova, P., Van Hout, R. and Scharenborg, O. "Lexically guided perceptual learning in non-native listening", Biling. Lang. Cogn., 19: 914-920, 2016.