A CRITERIAL INTERLOCUTOR TALLY FOR SUCCESSFUL TALKER ADAPTATION?

Anne Cutler, L. Ann Burchfield and Mark Antoniou

The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, NSW, Australia
a.cutler@westernsydney.edu.au, ann@u.northwestern.edu, m.antoniou@westernsydney.edu.au

ABSTRACT

Part of the remarkable efficiency of listening is accommodation to unfamiliar talkers' specific pronunciations by retuning of phonemic inter-category boundaries. Such retuning occurs in second (L2) as well as first language (L1); however, recent research with emigrés revealed successful adaptation in the environmental L2 but, unprecedentedly, not in L1 despite continuing L1 use. A possible explanation involving relative exposure to novel talkers is here tested in heritage language users with Mandarin as family L1 and English as environmental language. In English, exposure to an ambiguous sound in disambiguating word contexts prompted the expected adjustment of phonemic boundaries in subsequent categorisation. However, no adjustment occurred in Mandarin, again despite regular use. Participants reported highly asymmetric interlocutor counts in the two languages. We conclude that successful retuning ability requires regular exposure to novel talkers in the language in question, a criterion not met for the emigrés' or for these heritage users' L1.

Keywords: talker adaptation, talker familiarity, L1, L2, heritage language.

1. INTRODUCTION

One of the most important features of human speech perception is its flexibility; talkers who have never previously been heard are understood on the first encounter. In the past decade and a half, this adaptation process has been extensively investigated using a paradigm in which listeners hear ambiguous phonetic forms, which can be disambiguated by reference to existing knowledge (see [1] for a review). The first use of this paradigm [2] established that exposure to just 20 exemplars of a deviant phonemic form induces retuning of the boundaries of the phoneme in question for that speaker – as long as the deviant form is heard in real-word contexts so that it can be assigned to a phonemic category. Thus an ambiguous sound between [s] and [f] will be judged to be [s] if heard in words like *loose*, to be [f] if heard in words like *cliff*, but will remain ambiguous if heard in nonwords such as *liff* or *liss*, because these will not provide the necessary lexical frame of reference to choose one potential category over the other.

The learning offers a path for adaptation on first exposure to speech from a new talker, because it generalises to other words with the same phoneme, it occurs with many types of phoneme, in varying positions in the word, and it is long-lasting, while its initiation is notably rapid: fewer than 20 exemplars will also produce retuning (for all details see [1]). It has been documented for, and is presumed to be used by, listeners from childhood to old age [3,4].

Importantly for the present work, this learning has been observed in many languages, both European and non-European. It also holds for lexically distinctive non-segmental speech sounds; thus in Mandarin, a lexical tone ambiguous between Mandarin tones 1 and 2 induced adjustment [5] similar to that seen for ambiguous phonemes in the same language [6]). Furthermore, the learning can be successfully applied not only in the L1 but also in L2 [7-10].

One study of such learning in both the L1 and L2 of the same listeners led to a surprising result. These listeners were emigrés, with Dutch as L1, but fluent English as L2 due to longterm residence (on average, 22 years) in an English-language environment. In the L2, these listeners showed significant adaptation; but in their L1, no significant learning appeared. This was particularly unexpected in that they reported, without exception, continued regular use of their L1 [9].

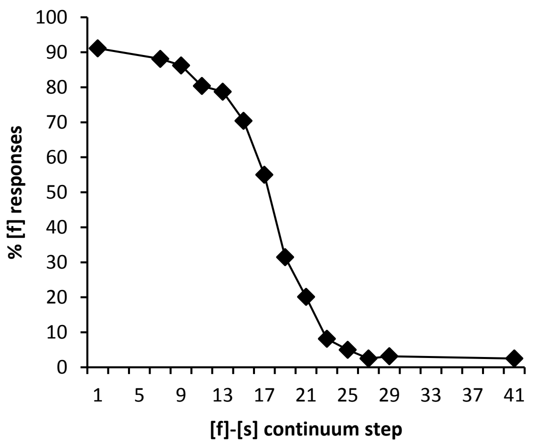
This unexpected finding could perhaps reflect the participants' higher age than in other L2 studies (though as noted, age is not a relevant factor in the case of the L1). It could also result from some aspect of the emigré situation. But also, and interestingly, it could indicate that talker adaptation mechanisms (in any language, L1 or L2) require regular practice for optimal operation. In such a case, regular use of a language only with long-known family members may need no adaptation, and hence the practice criterion may not be met.

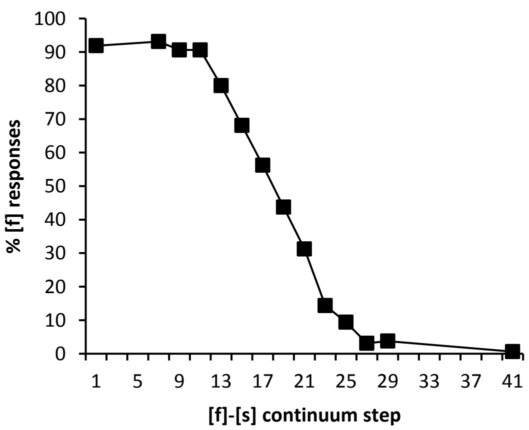
Emigrés are not the only bilingual user group for which there is typically a considerable asymmetry of interlocutor set size for the users' two languages; the same often occurs for heritage language users. In the heritage case, the home L1 is typically used by and with older family members, while younger members tend to use the environmental language at school, at work, and outside the family. In this study, we test the interlocutor tally account of the emigré results [9] with heritage language users.

2. PILOT EXPERIMENTS

To select an [f]-[s] ambiguous sound for use in the perceptual learning exposure phase, we conducted for each language a pilot experiment in which a 41-step continuum was created. The consonant-vowel (CV) syllables [fu], [su], and [θu] were produced by female native speakers of Mandarin and of English. The fricatives were excised from [fu] and [su] and then mixed to produce 41 equidistant steps ranging from 100% [f], 0% [s] to 0% [f], 100% [s]. These fricatives were then spliced onto the vowel [u] from [θu] (to eliminate the possibility of coarticulatory cues within the vowel biasing listeners to categorise the ambiguous fricative as [f] or [s]). A selection of 14 steps from the [f]-[s] continuum served as pilot stimuli: 1 ([f]), 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29 and 41 ([s]). Native listeners of the relevant language categorised 10 occurrences of each of these 14 steps; in each language, step 17 proved the most ambiguous token (see Fig. 1), and was thus used in the ambiguous stimuli for the exposure phase.

Figure 1: Pilot experiments: percentage of [f] responses for Mandarin (above) and English (below). For each language, a smooth categorical function was observed. Each step 17 was chosen as the ambiguous sound for use in the exposure task.





3. PERCEPTUAL LEARNING EXPERIMENTS

3.1 Method

The perceptual learning experiments consisted of an exposure phase in which participants heard spoken items and decided for each whether it was a real word or a non-word, and a test phase in which they heard CV syllables and classified each C as [f] or [s]. Both English and Mandarin materials sets had previously induced successful perceptual learning in L1 [11].

3.1.1 Participants

Participants in the perceptual learning experiments were 25 Mandarin-English early bilinguals (mean age 22.2, *SD* = 4.9), none of whom reported any vision, hearing, or language impairments. In comparison to the emigré listeners in [9], these participants were much younger, and were not emigrés (all were born in and still resided in Australia). Moreover, their two languages, Mandarin and English, are phonologically dissimilar, unlike English vs. Dutch in the case of [9].

3.1.2 Exposure phase: Training materials

The lexical decision task in each language involved 200 disyllabic items, half being words and half non-words. Of the words, 40 were training items and 60 fillers. Half of the training items were [f]-words ([f] as the first phoneme of the second syllable, e.g., *bu4fa3* 'illegal'; *traffic*), and half were s-words ([s] in the same medial position: e.g., *kuan1song1* 'loose'; *insane*). Substituting [f] for [s] in any training item formed 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). The mean frequency for English words was 4.3 and 4.1 respectively for f-words and s-words (computed from the SUBTLEX Zipf scale).

Two versions of each training item were created: one natural, and one modified in that the critical word-medial fricative was replaced by the ambiguous sound (step 17 from the pilot experiment continuum in each case; hereafter [?]). Neither [f] nor [s], nor the similar sounds [ʂ], [ɕ], [ts], and [tʃ] occurred in other words, or in the nonwords or fillers. Four presentation orders were created, in which no more than four words or non-words occurred in a row, and [?] did not occur within the first 12 trials.

3.1.3 Test phase: Materials

Test materials were five steps (7, 13, 17, 21, 27) of the /fu/-/su/ continuum; they were presented 30 times in random order (150 trials in total). All pilot, training and test materials were produced by the same speaker.

3.1.4 Procedure

Each participant completed the perceptual learning experimental tasks in both Mandarin and English. These sessions occurred 2-3 weeks apart, with order of language balanced as far as possible (12/13); the same held for assignment to the [f]-trained group, for whom [?] replaced all occurrences of [f], versus to the [s]-trained group, for whom all occurrences of [s] were replaced. Each participant was trained on one sound per language; the trained category was the same across languages for half the group, different for the other half. In each lexical decision task the participants decided for each stimulus item whether it was a real word or a non-word. In the test task, which followed completion of the lexical decision task, they categorised the 150 CV tokens as either [fu] or [su]. At the end of their initial experimental session, all participants completed a language use questionnaire.

3.2 Results

3.2.1 Lexical decision results

In no analysis was the factor session order significant and to save space it has been omitted from this report. All participants scored above 55% in lexical decision in both languages. In Mandarin, the f- and s-trained groups did not differ (77.6% vs. 78.3%, $t(23) = -.14$, $p = .891$). Both groups correctly rejected Mandarin nonwords (78.6% and 81.5%, respectively).

Percentages of ambiguous items accepted as words in the lexical-decision task were calculated for each participant. The f-trained group accepted 64.5% of the f-ambiguous items and the s-trained group accepted 60.4% of the s-ambiguous items as words. We also tested whether the rate at which ambiguous items were accepted as words altered over the course of the lexical decision task, by dividing training trials into four quartiles. Consistent with prior evidence that perceptual retuning requires only minimal exposure, we found no acceptance rate difference in ambiguous items across training (f-trained group across quartiles 1 to 4: 60.0%, 58.2%, 78.2%, 61.8%; s-trained group across quartiles 1 to 4: 60.0%, 56.7%, 66.7%, 58.3%).

In English, both f- and s-trained groups correctly rejected English nonwords (87.0% and 74.5%, respectively). The groups differed however in their lexical decisions (overall respectively 91.0% vs. 80.6%; $t(23) = 4.45$, $p < .001$). The f-trained group accepted 96.9% of the 20 f-ambiguous items whereas the s-trained group only accepted 62.5% of the s-ambiguous items as words. This is a sign of the "f-bias" frequently observed in [f]-[s] retuning (see, e.g., [2]), which is held to stem from the relative salience of acoustic cues to [s] (higher) vs. [f] (lower). A quartile analysis here showed consistent high rates of acceptance for the f-trained group (Q1 = 95.4%, Q2 = 98.5%, Q3 = 96.9%, Q4 = 96.9%). Inspection of the s-trained group revealed that three subjects had rejected the majority of s-ambiguous items (of 20 items, one rejected 17, another 18, the third all 20). With these subjects excluded, the s-trained group quartile scores were Q1 = 60.0%, Q2 = 86.7%, Q3 = 88.9%, Q4 = 86.7%, indicating learning in early trials.

3.2.2 Categorisation results

Figure 2: Percentage of [f] responses by Mandarin-English early bilinguals to a Mandarin [fu]-[su] continuum as a function of training group.

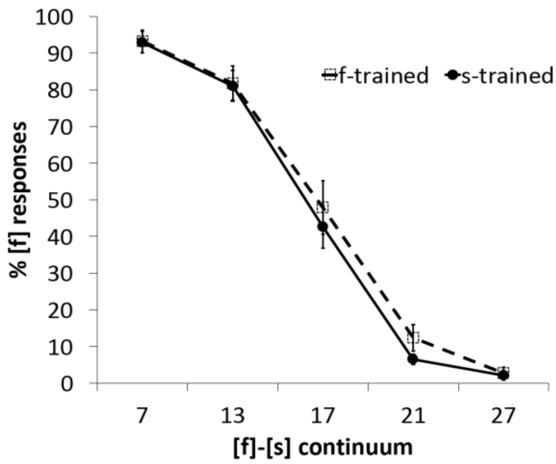
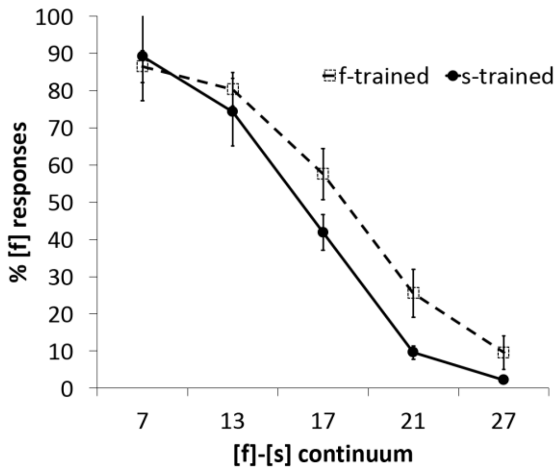


Figure 3: Percentage of [f] responses by Mandarin-English early bilinguals to an English [fu]-[su] continuum as a function of training group.



Perceptual learning was assessed by a 2 × 5 ANOVA with the between-subjects factor of training group ([f]-trained vs. [s]-trained) and the within-subjects factor of step (7, 13, 17, 21, 27) separately for each language. The Mandarin study showed no evidence of perceptual learning (see Fig. 2). There was a significant effect of continuum step, $F(4, 92) = 314.5$,

$p < .001$, $\eta_p^2 = .937$, but no significant effect of group, $F(1, 23) = 0.5$, $p = .507$, $\eta_p^2 = .021$, and no interaction, $F(4, 92) = 0.3$, $p = .845$, $\eta_p^2 = .016$.

In English, in contrast, the results clearly showed separation between training groups (see Fig. 3). Here there was a significant main effect of step, $F(4, 92) = 145.7$, $p < .001$, $\eta_p^2 = .864$, a marginal effect of group, $F(1, 23) = 3.6$, $p = .071$, $\eta_p^2 = .135$, and no significant interaction, $F(4, 92) = 1.3$, $p = .286$, $\eta_p^2 = .052$. Removal from the s-group of the three subjects who showed no acceptance of [?] as [s] resulted in the main effect of group becoming significant ($p = .03$) while the step effect and interaction did not change.

3.2.3 Questionnaire results

The questionnaire we used contained 126 items in all, addressing background, family, education and current usage for each language across a wide range of situations and purposes. For the present analyses we draw on the interactional and usage questions only (11 items excluding those on interaction with partners and pets, which proved not relevant for our sample). These were: dominant language; most comfortable language to speak; friends' native language; rated proficiency; language used with family; language used with non-family; language used at work (N.B. some questions were rated separately by language). As expected, the two languages showed sharp response differences. English had 96% of the 25 participants reporting dominance, 91.7% talking it always or mostly with acquaintances and friends, and 86.3% using it more at work. Mandarin, in contrast, was ranked significantly higher than English on just one point: 83.3% reported Mandarin as sole or usual language used with family.

We examined these findings for correlation with a measure of learning devised by [4] based on the test data, namely mean training-consistent responses to the most ambiguous categorisation steps (13, 17, 21). This measure did not differ across the f- and s-trained groups on the Mandarin test, $t(23) = 0.69$, $p = .248$, but on the English test the f-group was slightly (but significantly) more consistent: $t(23) = 1.99$, $p = .029$.

Mandarin exposure task results correlated with the responses to "To what extent do you use Mandarin with relatives?", $r(24) = .588$, $p = .002$, and inversely with the same question re English, $r(24) = -.474$, $p = .011$ (the higher the learning scores, the more Mandarin and the less English). English exposure task results correlated with responses to "What is the native language of the majority of your friends and acquaintances who live here?", $r(24) = .635$, $p = .001$ (higher scores, more English choices). Mandarin test results correlated negatively with responses to "To what extent do you use English at work?", $r(24) = -.444$, $p = .017$ (higher learning, less English). English test results correlated with responses to "Do you feel more comfortable speaking Mandarin, English or both?", $r(24) = .433$, $p = .020$ (higher learning, more comfortable with English). Our small sample size as well as the high degree of skew in the questionnaire responses make it difficult to draw strong conclusions regarding these correlations; we note, however, that all significant findings were in the expected direction.

4. CONCLUSION

This study has supported the prediction that success in talker adaptation by phoneme category retuning is likely to be greater when the language in question is spoken with a larger number of different interlocutors. This prediction was motivated by the prior finding [9] that longterm emigrés to an English-speaking country showed significant retuning in their L2 English, but no significant learning in their L1, despite the fact that the L1 remained in regular use. Based on the evidence that their L1 interlocutors were principally long-known family members, the prediction of [9] was that variability among interlocutors is essential for the maintenance of category retuning skill in a language.

The early-bilingual heritage language users in the present study also reported a consistent difference in the interlocutor tally per language; the language used with family was most often Mandarin, while English was the dominant language and the language most used in interactions outside the family. Note that again all our participants reported regularly using both languages and many reported that speaking each was equally comfortable. However, their perceptual learning success was greater in the language in which they talked with more different people: English. In the family L1, Mandarin, no significant learning appeared.

Alternative explanations of the [9] result (L1-L2 similarity, participant age, or specificity to the emigré situation), cannot account for the findings of our study, which differed from [9] in all these respects. It also cannot be the case that an adaptation asymmetry only occurs if the L2 is acquired later than the L1.

Conversing with multiple interlocutors thus seems to keep listeners' mechanisms of talker adaptation in good working order. This is not the only way in which L2 skills benefit from talker variability; perception of difficult L2 phoneme contrasts also benefits from training on productions by multiple talkers [12,13]. It is also not the only way in which patterns of language use adversely affect language skill, from fossilisation in L2 [14] to attrition of L1 [15]. We predict, though, that the L1 disadvantage we have shown, supporting the conjecture in [9], is likely not to be permanent, but rather, reparable by means of experiential alterations. That however remains a question for future research.

5. ACKNOWLEDGEMENT

This research was supported by the Australian Research Council (grant DP 140104389).

6. REFERENCES

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

[2] Norris, D., McQueen, J. M., Cutler, A. 2003. Perceptual learning in speech. *Cogn. Psychol.* 47, 204–238.

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

[4] Scharenborg, O., Janse, E. 2013. Comparing lexically guided perceptual learning in younger and older listeners. *Atten. Percept. Psychophys.*, 75, 525–536.

[5] Mitterer, H., Chen, Y., Zhou, X. 2010. Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm, *Cogn. Sci*. 35, 184–197.

[6] Burchfield, L. A., Luk, S. H. K., Antoniou, M., Cutler, A. 2017. Lexically guided perceptual learning in Mandarin Chinese. *Proc. Interspeech 2017* Stockholm, 576–580.

[7] Reinisch, E., Weber, A., Mitterer, H. 2013. Listeners retune phoneme categories across languages, *J. Exp. Psychol. Hum. Percept. Perform.* 39, 75–86.

[8] Schuhmann, K. S. 2014. *Perceptual learning in second language learners*. PhD thesis, State University of New York at Stony Brook.

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

[10] Drozdova, P., van Hout, R., Scharenborg, O. 2016. Lexically-guided perceptual learning in non-native listening, *Biling. Lang. Cogn.* 19, 914–920.

[11] Cutler, A., Burchfield, L.A., Antoniou, M. 2018. Factors affecting talker adaptation in a second language. *Proc. 17th Austral. Internat. Conf. Speech Sci. and Technol.*, Sydney, 33-36.

[12] Lively, S. E., Logan, J. S., Pisoni, D. B. 1993. Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *J. Acoust. Soc. Am.* 94, 1242–1255.

[13] Wang, Y., Spence, M., Jongman, A., Sereno, J. 1999. Training American listeners to perceive Mandarin tones. *J. Acoust. Soc. Am.* 106, 3649–3658.

[14] Long, M. H. 2003. Stabilization and fossilization in interlanguage development. In: Doughty, C., Long, M. H. (eds), *The Handbook of Second Language Acquisition*. New York: Blackwell-Wiley, 487-534.

[15] Schmid, M. S. 2013. *First Language Attrition*. Amsterdam: John Benjamins.