Is retroflexion a stable cue for distributional learning for speech sounds across languages? Learning for some bilingual adults, but not generalisable to a wider population in a well powered pre-registered study

Hannah L Goh [Corresp., 1] , **Luca Onnis** [2] , **Suzy J Styles** [3, 4]

[1] Interdisciplinary Graduate Programme, Nanyang Technological University, Singapore, Singapore

[2] Center for Multilingualism in Society across the Lifespan, University of Oslo, Oslo, Norway

[3] School of Social Sciences, Nanyang Technological University, Singapore, Singapore

[4] The Centre for Research and Development in Learning (CRADLE), Nanyang Technological University, Singapore, Singapore, Singapore

Corresponding Author: Hannah L Goh
Email address: hannahle001@ntu.edu.sg

Bilinguals are widely reported to have certain kinds of cognitive advantages, including language learning advantages. One possible pathway is a language-specific transfer effect, whereby sensitivity to structural regularities in *known languages* can be brought to *to-be-acquired languages* that share particular features. Here we tested for transfer of a specific linguistic property, sensitivity to retroflexion as contrastive phonemic feature. We designed a task for bilinguals with homogeneous language exposure (i.e., bilingual in the same languages) and heterogeneous feature representation (i.e., differing levels of proficiency). Hindi and Mandarin Chinese both have retroflexion in phoneme contrasts (Hindi: stop consonants, Mandarin: sibilants). In a preregistered study, we conducted a statistical learning task for the Hindi dental-retroflex stop contrast with a group of early parallel English-Mandarin bilinguals, who varied in their Mandarin understanding levels. We based the target sample size on power analysis of a pilot study with a Bayesian stop-rule after minimum threshold. Contrary to the pilot study ($N = 15$), the main study ($N = 50$) did not find evidence for a learning effect, nor language-experience variance within the group. This finding suggests that statistical effects for the feature in question may be more fragile than commonly assumed, and may be evident in only a small subsample of the general population (as in our pilot). These stimuli have previously shown learning effects in children, so an additional possibility is that neural commitment to adults' languages prevents learning of the fine-stimulus contrast in question for this adult population.

1 **Is retroflexion a stable cue for distributional learning for speech sounds across languages? Learning for**

2   **some bilingual adults, but not generalisable to a wider population in a well powered pre-registered**

3   **study**

5 Hannah L. Goh[1], Luca Onnis[2] & Suzy J Styles[3,4]

6 [1]Interdisciplinary Graduate Programme, Nanyang Technological University, Singapore

7 [2]Center for Multilingualism in Society Across the Lifespan, University of Oslo, Norway

8 [3]Psychology, School of Social Sciences, Nanyang Technological University, Singapore

9 [4]The Centre for Research and Development in Learning, Nanyang Technological University, Singapore

11 Contact Authors hannahle001@e.ntu.edu.sg

13 **Authorship statement**

14 HLG, LO, & SJS designed the study. HLG conducted the study, and analysed the data with guidance from

15 SJS. SJS secured funding for the project. HLG, LO, and SJS wrote the paper. The authors confirm they have

16 no conflicts of interest.

18 **Acknowledgements**

19 This research was supported by the following funding sources: Singapore's National Research Foundation

20 under the Science of Learning (NRF2016-SOL002-011), the Centre for Research and Development in

21 Learning, Nanyang Technological University (JHU IO 90071537). The authors would like to thank Asst Prof

22 Scott Reid Moisik and Prof Annabel Chen for their insightful feedback on this project, as well as Asst Prof

23 Maaike Vandermosten and Assoc Prof Narly Golestani for sharing their stimuli and procedures with us for

24 our adaptation of their study.

27 **Abstract**

28   Bilinguals are widely reported to have certain kinds of cognitive advantages, including language

29   learning advantages. One possible pathway is a language-specific transfer effect, whereby sensitivity

30   to structural regularities in *known languages* can be brought to *to-be-acquired languages* that share

31   particular features. Here we tested for transfer of a specific linguistic property, sensitivity to

32   retroflexion as contrastive phonemic feature. We designed a task for bilinguals with homogeneous

33   language exposure (i.e., bilingual in the same languages) and heterogeneous feature representation

34   (i.e., differing levels of proficiency). Hindi and Mandarin Chinese both have retroflexion in phoneme

35   contrasts (Hindi: stop consonants, Mandarin: sibilants). In a preregistered study, we conducted a

36   statistical learning task for the Hindi dental-retroflex stop contrast with a group of early parallel

37   English-Mandarin bilinguals, who varied in their Mandarin understanding levels. We based the target

38   sample size on power analysis of a pilot study with a Bayesian stop-rule after minimum threshold.

39   Contrary to the pilot study ($N = 15$), the main study ($N = 50$) did not find evidence for a learning effect,

40   nor language-experience variance within the group. This finding suggests that statistical learning

41   effects for the feature in question may be more fragile than commonly assumed, and may be evident

42   in only a small subsample of the general population (as in our pilot). These stimuli have previously

43   shown learning effects in children, so an additional possibility is that neural commitment to adults'

44   languages prevents learning of the fine-stimulus contrast in question for this adult population.

45   **Keywords:** bilingual learning advantage, statistical learning, retroflexion

46

47   **Introduction**

48   Bilinguals tend to show a learning advantage for acquiring novel language structures. For example, bi- and

49   multilinguals show enhanced learning for grammatical structures (Kemp, 2007; Kovács & Mehler, 2009),

50   acquiring new vocabulary (Kaushanskaya & Marian 2009; Yoshida et al., 2011), and developing sensitivity

51   to phoneme contrasts (Antoniou et al., 2015; Singh et al., 2018; Wang & Saffran, 2014). In addition,

52   bilinguals may be better at learning novel language-like structures in artificial grammar learning tasks

53   (Onnis et al., 2018). While the exact mechanisms underlying this advantage continue to be explored, we

54   have identified three main overarching hypotheses in existing literature. We refer to these hypotheses as

55   domain-general advantages, general linguistic advantages, and specific linguistic advantages.

56   Firstly, *domain-general* accounts propose that acquiring multiple languages may lead to the development

57   of cognitive enhancements that boost learning for novel structures – including linguistic structures.

58   Support for a *general bilingual advantage* comes from cognitive studies comparing groups with different

59   language backgrounds. For instance, studies have found that bilinguals may have significantly better

60   executive control and selective attention than monolinguals (e.g., Bialystock & Craik, 2010; Carlson &

61   Meltzoff, 2008). However, it is unclear whether such domain-general cognitive advantages would

62   contribute to the specific case of language learning. Some studies have investigated correlations between

63   executive control and language-related learning. Yoshida et al. (2011) reported a bilingual advantage and

64   correlations between the performance of 3-year-olds on language and executive function tasks. However,

65   Bartolotti et al. (2011), found mixed evidence, with bilingual experience influencing learning of two

66   conflicting artificial languages only in low interference conditions, and inhibitory control influencing

67   learning in high interference conditions. Hence, relationships between bilingualism and cognitive

68   functions may not necessarily be linear; instead, bilingual learning advantages may merely be mediated

69   by general cognitive functions. Furthermore, the question of whether bilingualism actually affords this

70   kind of general cognitive advantage remains hotly debated. A recent review by Antoniou (2019) found

71   that studies often disagree whether bilinguals show improved cognitive functions, and whether some

72   groups of bilinguals demonstrate enhanced cognitive functions (e.g., children and the elderly) while others

73   (e.g., young adults) do not. In current literature, therefore, it is difficult to establish how much a domain

74   general bilingual advantage contributes to the learning of new language structures.

75   Secondly, *general linguistic* accounts focussing on the linguistic nature of the learning advantage

76   propose that bilingual learning advantages are grounded in the linguistic experience of managing and

77   making sense of multiple languages. By extension, acquiring new linguistic patterns and structures would

78   therefore be easier for individuals who already have more than one linguistic pattern in their language

79   repertoire. For instance, Tremblay and Sabourin (2012) trained adult participants on a novel Hindi dental-

80   retroflex contrast and found that bilinguals and multilinguals showed larger learning effects and better

81   transfer effects to a novel but related phoneme contrast than monolinguals. Further support for the idea

82   that linguistic structural overlap might mediate bilingual learning advantages can be seen in a study by

83   Enomoto (1994) who found general bilingual advantages in sensitivity to a Japanese geminate contrast

84   were stronger, especially when specific phonetic features overlapped with the bilinguals' linguistic

85   experiences.

86   Therefore, while there is some evidence to suggest there is a *general linguistic advantage* for bilinguals,

87   it may be the case that learning advantages are *specific to properties of the languages in question*, and

88   may rely on the degree of overlap between the structure of the languages in a speaker's repertoire and

89   the to-be-acquired linguistic content. For instance, a three-year longitudinal study by Kopečková (2016)

90   found that individuals who were familiar with the alveolar /r/ phoneme in at least one of their existing

91   languages (e.g., German-Croatian, and German-Russian bilinguals) were better at learning to produce the

92   Spanish alveolar trill /r/, as compared to bilinguals without similar linguistic overlaps. Similarly, Antoniou

93   et al. (2015) tested English monolinguals, Mandarin-English bilinguals, and Korean-English bilinguals on

94   their ability to learn a novel place-of-articulation contrast based on the phonology of Mandarin, and an

95   artificial lenition contrast based on the phonology of Korean. They found that while both bilingual groups

96   were better than English monolinguals at learning the Mandarin-based place of articulation contrast, only

97   the Korean-speaking bilinguals showed a learning advantage for the more difficult Korean-based lenition

98   contrast. This finding supports the hypothesis that learning advantages may be specific to linguistic

99   overlaps between existing linguistic features and no-be-learned stimulus contrasts. There is evidence to

100   support that fact that *language-specific bilingual advantages* may be long-lasting and can be present even

101   if exposure to a language only occurred for a short period early in life. For instance, Werker (1986)

102   demonstrated that adults who had only been exposed to Hindi up to the age of two were significantly

103   better at discriminating a difficult Hindi dental-retroflex contrast as compared to monolingual and

104   bilingual adult participants who did not have any experience with Hindi across the lifetime. Similar

105   forgotten-language results were also found by Singh and Seet (2019), who found that adults who had

106   been raised by Hokkien speaking caregivers up to the age of three were significantly better at learning

107   Hokkien tonal contrasts than adults who had only been raised by English speaking caregivers. Likewise,

108   Oh et al. (2009) found that Korean-Americans adopted from Korea in early childhood were significantly

109   better at learning Korean aspirated and lenition phonemes compared to individuals with no early

110   childhood Korean exposure. The impact of language specific learning effects may also extend to the way

111   in which novel grammar is learned. Onnis & Thiessen (2013) exposed Korean and English adults to a

112   statistical learning task, in which forward and backward probabilities between adjacent pseudowords

113   generated two equally probable and orthogonal patterns in their order of presentation, one more similar

114   to Korean, and the other more similar to English. Their results showed that the participants' probabilistic

115   preferences for pseudoword order aligned with the word order pattern that was more similar to their L1

116   language, indicating that their learning was heavily influenced by their prior experience of linguistic

117   patterns. Moreover, while the Korean participants in the study had received extensive formal training in

118   English and lived in an English-speaking environment, they nevertheless exhibited statistical learning

119   biases congruent with their native Korean. These findings appear to show that the strength of *specific*

120   *bilingual linguistic advantages* may be predicted by the strength of representation that individuals have

121   with each of the languages in their repertoire. On the whole, these studies demonstrate that shared

122   phonetic and structural features appear to be powerful predictors of linguistic learning advantages, thus

123   supporting a *specific bilingual linguistic advantage* based on familiarity of shared features or regularities.

124   To explore the language specific overlap hypothesis, we decided to examine whether a language specific

125   learning effect could be found in an auditory statistical learning task for English-Mandarin bilingual adults,

126   if the to-be-learned contrast included a phonetic feature that has contrastive functions in Mandarin. For

127   auditory targets, we selected a Hindi dental-retroflex phoneme contrast identical to that of studies carried

128   out by Golestani et al., (2002), Golestani & Zatorre (2009), and Vandermosten et al. (2018) on children

129   and adults. The phonetic feature of retroflexion is shared between Hindi and Mandarin Chinese, albeit in

130   different phonemes (Hindi: stops; Mandarin: sibilants), thereby providing linguistic overlap with the

131   phonology of one of our bilingual participants' languages. Hence, while English-Mandarin bilinguals will

132   have familiarity with an alveolar-retroflex contrast in sibilants such as /s/ vs. /ʂ/, they will be unfamiliar

133   with the Hindi dental-retroflex stop contrast of /d̪/ vs. /ɖ/. Furthermore, our participants -Singapore's

134   English-Mandarin bilinguals- vary in their Mandarin proficiency, exposure, and use. As a result, we expect

135   that there will be individual differences in Singapore's English-Mandarin bilinguals' familiarity with the

136   retroflex feature of Standard Mandarin (i.e., the Beijing accented variety), allowing us to explore the

137   question of whether the strength of a potential language specific learning effect might be linked to

138   strength of representation that individuals have with the overlapping feature in question.

139   We chose to base our study on a methodological adaptation of Vandermosten et al. (2018)'s passive

140   listening statistical learning paradigm, adapted for adults. Statistical learning involves the process of

141   learning by extracting regularities from sensory input in the environment, and plays a critical role in the

142   process of early language acquisition, particularly in the learning of native-language phoneme contrast

143   boundaries. According to Werker, Yeung, and Yoshida (2012) the extraction of information from the

144   relative frequencies of meaningfully contrastive speech sounds heard early in life is crucial to the

145   formation of perceptual boundaries around unique native language phonetic categories. Examples of this

146   can be seen in infant studies that track the development of phoneme sensitivities that shift from a more

147   general ability to distinguish between a wide variety of phoneme contrasts, to becoming more aligned to

148   infants' native languages over the course of several months in early infancy (e.g., Werker & Tees, 1984).

149   The ability to learn new phoneme contrast categories through statistical learning continues throughout

150   childhood, as children have the ability to track statistical information about frequencies of exposure to

151   different sound tokens arranged on an acoustic continuum. For instance, when trained on a bimodal

152   distribution of speech sounds (with more tokens drawn from steps closer to the ends of an acoustic

153   continuum), children have been shown to develop a two-category perception of the continuum, behaving

154   as though the continuum consists of two main contrastive categories of speech sounds (e.g.,

155   Vandermosten et al., 2018). However, studies of passive statistical learning in adults differ in whether

156   they generate expected learning effects. Terry et al. (2015), for example, found that adult Australian

157   English speakers were unable to show significant learning effects following bimodal distributional training

158   on the Dutch /ɑ/-/a:/ phoneme contrast, even with the use of stimuli that enhanced the differences in

159   the vowel categories, and an event-related potential mismatch negativity study by Wanrooij et al. (2014)

160   found that Dutch-speaking adults could not discriminate between the English vowels /æ/ and /e/

161   following bimodal training on the vowel sounds. At the same time, Hayes-Harb (2007) also demonstrated

162   that while adults may be able to extract some degree of novel phoneme category information from

163   statistical learning, they appear to rely more heavily on lexical information in their learning of novel

164   contrasts. On the other hand, stronger evidence of statistical learning in adults has been found in some

165   studies. Chládková and Šimáčková (2021) found Czech speakers were able to learn a novel Czech-like

166   durational contrast following bimodal distributional training while Greek speakers who were not familiar

167   with this linguistic feature were not able to learn. Similar evidence of adults benefitting from statistical

168   learning can also be seen in a series of studies that show that Dutch-learning native Spanish speakers

169   living in the Netherlands show a significant improvement in their perception of the difficult Dutch /ɑ/-/aː/

170   vowel contrast following bimodal distributional training (Escudero et al., 2011; ; Escudero & Williams,

171   2014; Wanrooij et al., 2013). Critically, one key factor in the studies where adults *were* able to learn from

172   statistical information is that they all had some degree of overlap in their language repertoire (phonology,

173   word order) with the stimulus dimension that they were being trained on.

174   Hindi and Mandarin Chinese both have retroflexion in phoneme contrasts (Hindi: stop consonants,

175   Mandarin: sibilants). Based on existing evidence, we thus predicted English-Mandarin bilinguals adults

176   that would show a learning effect for a Hindi dental-retroflex contrast in a distributional learning paradigm

177   similar to that of Vandermosten et al. (2018). Therefore, we decided to base our exploration of the

178   language specific effect on Singapore English-Mandarin bilingual adults by determining if their familiarity

179   with the Mandarin retroflex contrast could be related to their statistical learning of retroflexion in Hindi

180   stop consonants. We first conducted a pilot study, and then a full study powered on the basis of pilot

181   results.

182   **Pilot Study**

183   In a pilot study, we tested if adult participants would be able to learn the Hindi dental-retroflex contrast

184   with a distributional learning paradigm. To begin with, all participants were tested on their ability to

185   discriminate the contrast. Following this pre-test, there was a training period in which participants were

186   exposed to sounds from the Hindi-dental retroflex continuum, with presentation frequencies

187   representing a bimodal frequency distribution. A post-test was used to establish whether learning had

188   occurred. The methodology of Vandermosten et al. (2018) was adapted in the following ways: Firstly, we

189   increased the number of sound tokens that participants heard during training, to maximise possible

190   learning effects, while maintaining the proportion frequency of each sound token such that this was

191   consistent with the original Vandermosten et al. (2018) study. Secondly, rather than comparing results

192   between learners and a control group trained on a unimodal distributional frequency, we used learners'

193   performance accuracy in pre-training as a baseline for comparison. Lastly, we conducted an unattended

194   training paradigm in which participants were told that they did not need to pay attention to the sounds,

195   rather than the original attended auditory task, to reduce working memory load.

196   **Methods**

197   **Participants.** 16 participants (10 female) were recruited from thestudent population of Nanyang

198   Technological University in exchange for course credits. Since colour was a key feature of the test stimuli,

199   one male participant was excluded due to colour blindness. The remaining 15 participants were aged

200   between 20-29 years old (*Median* = 22), were ethnically Chinese Singaporeans, and reported being fluent

201   in Mandarin and English, but not Hindi (see detailed Language ID profiles below). This study was approved

202   by the IRB of the host institution Nanyang Technological University (IRB-2019-01-034), and written

203   consent was obtained from all participants prior to the study.

204   **Stimuli.** Visual stimuli were two cartoon aliens (purple and orange), and a "transmission device" that

205   illustrated 'sending messages to aliens.' During the training phase, participants were also presented with

206   cartoons that were muted (selected clips from Series 2 of Mr Bean: The Animated Series; Atkinson et al.,

207   2003). Audio stimuli were identical to and taken from Vandermosten et al. (2018). As we did not create

208   these stimuli, details on the creation of the auditory tokens can be found in Golestani et al. (2022). As in

209   the original study, the stimulus continuum consisted of seven sound tokens 220ms in duration. Sound 1

210   represents a dental /d/, and Sound 7, a retroflex /ɖ/, synthesized in the original study to be close to the

211   category boundary for native speakers of Hindi. The original stimulus set also included intermediate audio

212   tokens (Sounds 2-6) synthesized at evenly spaced acoustic intervals between these two tokens, making a

213   continuum from the /d/ to the /ɖ/. Each consonant in the stimulus set was followed by the vowel /a/,

214   making a dental-retroflex da continuum of 7 sound tokens. All stimuli were presented on OpenSesame

215   (Mathôt et al., 2012). Code for the study is archived in the repository for this paper (Goh et al., 2020).

216   **Procedure**

217   **Familiarization Phase.** The goal of this phase was to introduce the participants to the two phoneme

218   categories, and to allow them to familiarise themselves with the sound-to-alien pairing. In an onscreen

219   task, participants were told that two aliens make different sounds. On each familiarisation trial,

220   participants were presented with an image of one of the aliens at a random position on the screen while

221   a sound was played. An orange alien was paired with Sound 1 (the dental /d/ end of the spectrum), and a

222   purple alien was paired with Sound 7 (the retroflex /ɖ/ end of the spectrum). Participants completed one

223   familiarisation block in which each alien was presented 10 times in alternating order. No response was

224   required during the familiarisation block.

225   **Practice Phase.** To ensure that participants were aware of the sound-to-alien pairing, they then completed

226   a practice block in which they were asked to 'send messages to aliens' using the pairing they had just

227   observed. On each trial, participants saw both aliens (See Figure 1) and one of the previously familiarized

228   stimulus tokens was played. Participants were asked to select an alien by pressing the key below it.

229   Keypress responses were followed by an animated 'beam' from the 'transmission device' to the selected

230   alien, along with a 'transmission sound' (created by Freesound.org contributor Jagadamba; downloaded

231   from https://freesound.org/people/Jagadamba/sounds/253908/), and onscreen feedback. Each stimulus

232   token was played four times in random order over the course of the block.

233   **Pre-Training Baseline.** This phase was similar to the practice phase, with the exception that the

234   participants heard all 7 of the sounds on the continuum and were not given feedback following a response

235   made. To establish a baseline for each person, participants completed three test blocks in which the

236   stimuli for each trial were randomly selected from the full stimulus continuum. Each of the 7 sounds was

237   played twice in random order over the course of each test block. There was no response time limit on

238   each test trial. Between each test block, participants completed a mini practice block (containing only

239   Sounds 1 and 7) to maximise their chances of remembering the correct sound/alien correspondence.

240 **Training Phase.** In the training phase, participants were asked to watch a cartoon while audio stimuli

241 played. In this unattended paradigm, a total of 4,020 sound tokens were presented to participants in a

242 random order, with an 80ms ISI between each sound. To achieve a bimodal stimulus distribution, each of

243 the 7 sounds from the stimulus continuum was played according to the distribution in Figure 2. The

244 training phase lasted around 20 minutes in total.

245 **Post-Training Test.** After the unattended training phase, participants completed a second round of

246 testing, identical to the Pre-training Baseline. No breaks were given between any of the phases.

247 **Language ID Profile.** To capture variance in participants' individual language exposure, at the end of the

248 alien message task participants were asked to rate their proficiencies in all of the languages/dialects they

249 understand on a Likert scale from 1, "I understand a few words in this language", to 7, "I have native-level

250 understanding of this language". The full demographic and language questions are in the repository for

251 this paper (Goh et al., 2020).

252 **Analysis Plan.** We measured the number of times each participant selected the purple alien (Sound 7

253 retroflex) or the orange alien (Sound 1 dental) for each stimulus token and computed the proportion of

254 Sound 7 responses at each step on the stimulus continuum. Psychometric curves were fitted for each

255 participant, and slope values were calculated using the *quickpsy* function in R (Linares & López-Moliner,

256 2016; R Core Team, 2020). Steeper slopes (i.e., higher slope values) represent greater sensitivity to

257 differences between bimodally distributed sounds on the dental-retroflex continuum, while flat slopes

258 represent chance responding. In order to compare pre-training baseline to post-training test, slopes were

259 computed separately for the pre- and post-training test blocks. As the sample size was small, a non-

260 parametric Wilcoxon signed-rank test was carried out to test for difference in the slope values in the pre-

261 and post-training test blocks.

262   In addition, in order to check whether any changes in slope value had occurred prior to the bimodal

263 exposure, we tested whether participants showed any differences between their first and third pre-

264   training test blocks. Individual slope values were derived separately for the three pre-training test blocks,

265   and a non-parametric Wilcoxon signed rank test was carried out. All analyses were carried out with the

266   statistical software R (2020).

267   **Results**

268   **Language Understanding.** Participants' self-reported English understanding was high (Range: 5-7, M = 6.5,

269   SD = 0.6) and Mandarin was more variable (Range: 3-7, M = 5.0, SD = 1.3). None of the participants

270   reported understanding any Indian languages or other languages with retroflexed phonemes. (See

271   distribution of participants' self-rated English and Mandarin understanding ratings in Figure 3.)

272   **Distributional Learning Effect.** To check whether any learning had taken place prior to the bimodal

273   exposure training phase, a Wilcoxon signed-rank test compared the first and third pre-training blocks. This

274   analysis revealed no significant difference ($Z(15)$ = -.03, p = .98; pre-1: slope M = 0.99, SD = 3.80; pre-3:

275   slope M = 0.60, SD = 2.13).

276   The fitted psychometric curves for pre-training and post-training can be seen in Figures 4a and 4b,

277   where it is clear that many participants show a steeper curve in the post-training phase. The Wilcoxon

278   signed-rank test revealed a significant difference in slope values between pre-training and post-training

279   ($Z(15)$ = -2.61, p = 0.004; pre- training slope: M = 0.5, SD = 0.19; post- training slope: M = 0.27, SD =

280   0.33), indicating that learning had occurred during the unattended training phase.

281   **Discussion**

282   The pilot study demonstrated a statistically significant learning effect in a methodological adaptation of

283   Vandermosten et al. (2018), conducted with adults. This finding suggests that the method and the stimuli

284   are suitable for use with bilingual adults in the local context and would be able to detect distributional

285   statistical learning of a non-native Hindi dental-retroflex contrast among Singaporean English-Mandarin

286   bilinguals. Furthermore, since individuals in the pilot showed a range of differences between their pre-

287 and post- training slope scores (pre-test: min = -0.30, max = 0.45; post-test: min = -0.2, max = 0.82), we

288 observed that individuals differed in the strength of the learning effect. This means that the paradigm is

289 suitable for further interrogation of individual differences in performance. Consequently, we

290 preregistered a full-scale study based on these pilot findings. While our preregistration also included a

291 Mandarin alveolar-retroflex sibilant word identification task, we did not include it in this study and the

292 results of the task were not included in any following analyses as there was a lack of meaningful variance

293 in the results of the task as the majority of our participants performed at ceiling level. Instead, the results

294 of this task have been presented in our paper on perception and production of the Mandarin alveolar-

295 retroflex contrast in Singapore (Goh et al., 2022).

296

297 **Main Study**

298 Following the pilot study, we carried out a full-scale study to further explore the language specific

299 hypothesis. We predicted that higher self-rated understanding of Mandarin would correspond to greater

300 familiarity with the retroflex feature, as this feature is prevalent in the standard variety of Mandarin (i.e.,

301 the Beijing variety), which is the dominant variety taught to bilinguals in the Singapore school system. If

302 language specific exposure is a source of bilingual learning advantages, then we expected to see greater

303 learning in individuals with higher Mandarin understanding scores.

304 The full-scale study was identical to the pilot, with the addition of supplementary language scales

305 designed to identify individual differences in participants' familiarity with Mandarin. To identify a suitable

306 sample size, we conducted an a priori power analysis in G*Power 3.1.9.7. This analysis revealed a

307 minimum sample size of 30 would be required to observe an effect size equal to the size observed in the

308 pilot (dz = 0.626) at an alpha level of 0.05 with a power of 1-β = 0.95. We increased the minimum sample

309 to N = 50 to ensure we had a minimum sample size suitable for an exploratory factor analysis (EFA) of

310 language factors (deWinter et al., 2009). In addition to this minimum sample size, we preregistered a data

311   collection 'stop rule' according to a Bayesian significance test on statistical learning effects to determine

312   if the results were substantially supportive of either H1 ($BF > 3$), or H0 ($BF < 0.33$) (Dienes, 2014). The prior

313   of 0.17 was determined using the size of the measured difference between pre- and post-training slope

314   values in the pilot study. The hypotheses and analysis plan were preregistered on the Open Science

315   Framework (Goh et al., 2020). The procedure was reviewed and approved by the ethics board of the host

316   institution Nanyang Technological University (IRB-2019-01-034).

317   **Methods**

318   **Participants.** In total, 56 participants (42 female) were recruited from the University student population

319   in exchange for course credits or payment ($15) following completion of the study. Three male and three

320   female participants were excluded from analysis, due to either a failure to follow instructions, or not being

321   from the target demographic. The remaining sample of 50 participants were aged between 18 – 25 years

322   of age (M = 20.3, SD = 1.7), all of whom were ethnically Chinese Singaporeans. Written consent was

323   obtained from all participants prior to the study.

324   **Procedure.** The procedure for the distributional learning and language understanding data collection

325   portions of the study were identical to the pilot.

326   **Results**

327   **Language ID Profiles.** On average, our participants' self-reported Mandarin understanding scores were

328   variable (Median = 6, Range: 3-7), while their English understanding was high (Median = 7, Range: 5-7),

329   indicating that they would have had been able to understand the experiment instructions (See Figure 5

330   for visualisation of English and Mandarin proficiencies in participants). Two participants also reported

331   understanding a few words in Tamil, with both rating their Tamil understanding with a score of 1. Two

332   variables were obtained from participants' self-rated language understanding scales for inclusion in

333   further analyses. The Mandarin understanding score of each participant was obtained directly from the

334   number rating that each participant gave for Mandarin understanding on the Language ID rating scale.

335   We also checked our participants' English understanding scores to ensure that all our participants had a

336   sufficient level of English understanding to fully understand the study instructions, all of which were

337   presented in English.

338   **Distributional learning.** To check whether exposure to the sound tokens heard during the pre-test blocks

339   lead to learning before the bimodal exposure training period, we conducted a preregistered control

340   analysis. No significant difference was observed in slope between the first and last pre-training blocks

341   (t(49) = 1.04; p = .30; pre-1: M = .12, SD = .44; pre-3: M = .05, SD = .47), showing no significant pre-training

342   learning. Given that no pre-training learning occurred, we proceeded to the main analysis. The fitted

343   psychometric curves for each participant at pre- and post-training are shown in Figures 6a and 6b. As the

344   slope values did not fit a normal distribution, a log transformation was applied to approximate Gaussian

345   distributions.

346   A multiple mixed models analysis was conducted on our participants' individual log-transformed slope

347   values, with the categorical fixed factor of test phase (pre-training, post-training) and the interaction

348   between test phase and self-reported Mandarin understanding rating, with random intercept of

349   participant and random by-participant slope for test phase. Bayes factors were calculated for each of the

350   p-values obtained (Bayes calculator created by Palif based on the 2008 Dienes Bayes calculator). Priors of

351   0.17 for the categorial fixed factor of test phase, and -0.29 for the interaction between test phase and

352   Mandarin understanding rating were used for the Bayes factors with both priors being obtained from the

353   pilot study.

354   Figure 6 shows the fitted psychometric functions for each individual before and after training. The

355   analysis revealed no significant main effect of test phase (pre-training, post-training) on slope values, and

356   the Bayes factor indicated substantial evidence in support of the null hypothesis that learning did not

357   occur following training (pre-test: M = .06, SD = .25; post-test: M = 0.04, SD = .27; t(1,48) = -.54; p = .59,

358     $B_{H(0, .17)}$ = .0001). Figure 7 shows the derived slope values pre- and post-training, along with self-reported

359   understanding of Mandarin Chinese. No significant interaction was observed between test phase and self-

360   reported Mandarin understanding scores ($t(1,48) = .31$; $p = .76$; $B_{H(0, -.29)} = .72$). Contrary to the results of

361   the pilot study, the main study showed no evidence of learning at the group level and no evidence that

362   individual Mandarin understanding influenced learning rate between individuals within the group, likely

363   due to the absence of an overall learning effect.

364

365   **Preregistered Exploratory Factor Analysis (EFA)**

366   To find out whether language background was related in a meaningful way to the microstructure of the

367   task – for example whether people with different kinds of language backgrounds showed learning effects

368   over different timescales – we conducted a preregistered EFA, using the *princomp* function (R Core Team,

369   2020). In order to conduct this EFA, we derived two additional language measures from the Language ID

370   questionnaire: the total number of languages understood by each participant, and the overall language

371   understanding scores of each participant. We also computed three additional learning microstructure

372   measures from the distributional learning task: 'no-training baseline', 'plasticity', and 'elasticity'. A total

373   of six factors were entered into the EFA: Mandarin understanding score, total number of languages

374   understood, overall language understanding score, no-training baseline, plasticity, and elasticity.

375   **Language measures.** Total number of languages understood of each participant was computed as the

376   total number of languages the participant rated as 1 or above on the language profile questionnaire. The

377   overall language understanding score of each participant was computed as the sum of all the numerical

378   ratings the participants gave for each of their language proficiencies. For example, a participant who rated

379   their English understanding as 6 and their Mandarin understanding as 3 would have an overall language

380   understanding score of 9. Participants reported understanding an average of 3.9 languages each (Range:

381   3-7, $SD = 1.3$), resulting in an average overall understanding score of 17.1 (Range: 10-31, $SD = 4.1$).

382   **Slope Microstructure.** Figure 8 shows a schematic of three possible phases in a learning trajectory, where

383   an individual may show differences in their response pattern at a micro-structural scale, even in the

384   absence of a global learning effect. *No-training baseline change* was designed to check if any participants

385   exhibited any sensitivity to the target phoneme category prior to exposure training. This was obtained by

386   deducting the slope values of the first pre-training test block from the third pre-training test block, with

387   positive values indicating an improvement in perception prior to the exposure training phase. *Plasticity*

388   was designed to find out whether learning was evident immediately following training, relative to the final

389   pre-trained perceptual state. Plasticity was computed by deducting the slope values of the third pre-

390   training test block from the first post-training test block, with positive values indicating an increase in

391   perceptual ability immediately following the exposure training phase. *Elasticity* was designed to check for

392   possible learning attrition over time in the post-test phase. Elasticity was computed by deducting the slope

393   values of the first post-training test block from the third post-training test block, with a negative value

394   indicating learning attrition. On average, participants showed a no-training baseline change of -.07 (*SD*

395   =.48, range: -1.26 – 0.75). On average, participants showed a plasticity change of .04 (*SD* = .60, range: -

396   1.65-1.33). Finally, on average, participants showed an elasticity change of 1.74 (*SD* = 14.0, range: -19.5-

397   95.4).

398   **Results.** The EFA revealed only one component that accounted for more variance than one variable (i.e.

399   20%). Component 1 revealed strong relationships between language factors. Participants who had more

400   languages in their repertoire, and had a higher overall language understanding score were less likely to

401   show a shift in perception prior to training. On the other hand, relationships between Mandarin

402   understanding and plasticity were much weaker. Participants who had lower Mandarin understanding

403   scores were moderately less likely to show a shift in perception following training – this could indicate

404   that there was a degree of interaction between Mandarin understanding ability, and distributional

405 learning effect for the Hindi dental-alveolar contrast. However,, this EFA did not provide us with any

406 further clarification on learning patterns of interest. See Table 1 for factor loadings on each component.

407 **Discussion**

408 While we did observe a distributional learning effect in a small-scale pilot study, our preregistered full-

409 scale study did not show evidence of a distributional learning effect for the Hindi dental/retroflex contrast

410 in a well-powered study of English-Mandarin bilingual adults. One likely reason for the difference is that

411 the learning effect in the pilot may have been driven by a small number of strong learners who are less

412 representative of the general population. Both the pilot and main study were drawn from the student

413 population of the host university, using the same recruitment pathway, so there is no reason to suspect

414 different recruitment biases between the two phases of the research. Rather it seems to represent a

415 random sampling effect in the small pilot sample. In the preregistered analysis plan, we used Bayesian

416 hypothesis testing to qualify our results, and determine the optimal time to stop recruitment following

417 the minimum sample size. Unlike p-values which only provide an estimate of the likelihood of a false

418 positive result, the Bayes factor estimates how much more supportive the data are of either the

419 experimental hypothesis or the null (Dienes, 2016; Lakens et al., 2018). The Bayes factor for the main

420 effect of phoneme contrast sensitivity (pre-training vs. post-training) confirmed that our data was

421 substantially more supportive of the null hypothesis as compared to the experimental hypothesis (1,000

422 times more supportive of the null), allowing us to have confidence in our conclusion that learning did not

423 occur. This difference between findings in a small pilot sample and a larger, preregistered sample is a

424 powerful demonstration of the value of preregistration in confirmatory hypothesis testing. In addition,

425 we demonstrate the value of Bayesian inference as a threshold for ceasing to collect data after substantial

426 evidence against the experimental hypothesis has been accumulated.

427   Our pre-registered stop rule was based on the results of the main effect of learning in the pilot study,

428 as a learning effect would first need to be obtained before we could look at the impact of Mandarin

429 understanding on any individual differences in the strength of the observed learning. At the sample size

430 consistent with the stop rule, the Bayes factor calculated for the interaction between Mandarin

431 understanding and learning was not substantial. However, we conducted a supplementary linear mixed

432 effect model on our pilot participants' individual log-transformed slope values to assess the interaction

433 effect between self-rated Mandarin understanding and learning effect in the pilot study (fixed factor if

434 test phase: pre vs. post; random intercept of participant and random by-participant slope for test phase).

435 This analysis revealed that there was no significant interaction effect between the two factors $t(1,13)$ = -

436 2.02, p = .06, ηp^2 = .24. A power analysis carried out on G*Power 3.1.9.7. revealed that a minimum sample

437 size of 44 should be sufficient to observe an interaction effect size equal to that observed in the pilot (ηp^2

438 = .24; f = .56) at an alpha level of 0.05 with a power of 1-β = 0.95. Since the main study was terminated at

439 50 participants, the power analysis confirms that the study was well-powered to observe a significant

440 interaction, if one had been evident.

441   The lack of a learning effect in the main study was somewhat surprising given the results of the pilot

442 and previous reports of learning with the same stimuli (Vandermosten et al. 2018). However, as the

443 Vandermosten study was conducted with Dutch children in Grade 3, the difference between study results

444 may lie in the ability of adults to pick up on fine-grained acoustic contrasts that children are able to detect.

445 Indeed, some studies have shown that there may be a shift in perceptual plasticity for speech-sounds,

446 with younger learners able to learn category structure from subtle distributional cues, and older learners

447 showing more reliance on phonetic and lexical cues (Hayes-Harb, 2007; Werker, 2018).

448   In addition, prior evidence has shown that adults are poorer at detecting unfamiliar phoneme contrasts

449 (Best & Strange, 1992; Tees & Werker, 1985). This could have meant that the dental-retroflex contrast

450 was too difficult for our adult participants to learn within a single distributional learning paradigm alone.

451 Some studies have suggested that learning is possible for adults if the paradigm includes exaggerated

452 phoneme contrast hyperarticulation during training, similar to that of infant-directed speech (Escudero et

453   al., 2011), or in multiple sessions of bimodal distributional training (Escudero & Williams, 2014). However,

454   some groups of adults are still unable to learn a completely novel phoneme contrast with bimodal training,

455   even when the acoustic difference in categories is emphasised during training (Terry et al., 2015). Notably,

456   the sound tokens used in the current study were synthesised to be very finely tuned in terms of the

457   acoustic differences between each sound, with the tokens differing only slightly from each other on the

458   third formant and central frequency of the initial burst (Golestani et al., 2002). However, differences in a

459   naturalistic Hindi dental-retroflex contrast are more complex, as Verma and Chawla (2003) found key

460   differences in not only the third formant, but also in the first, second, and fourth formants in their analysis

461   of the Hindi dental-retroflex contrast. These additional formant transition differences could play an

462   important role in emphasising differences between the phoneme categories, leading to the synthesised

463   continuum being harder to discriminate than typical exemplars of the naturalistic speech contrast. Hence

464   the particular tokens used in the current study may well be beyond the sensitivity of most adults to result

465   in learning in a passive exposure task.

466   Finally, while standard descriptions of both Mandarin Chinese and Hindi phonology include the retroflex

467   place of articulation, it is possible that the tongue positions differ. In particular, Mandarin Chinese

468   retroflexion may have less backwards curvature at the tongue tip than the retroflexion common to Hindi.

469   Indeed, studies on Mandarin articulation have found that some Mandarin speakers use a "bunched"

470   tongue position instead (Ou & Guo, 2014; Ladefoged & Maddieson, 1996; Luo, 2020). Moreover, some

471   studies have shown that speakers of "outer-circle" varieties of Mandarin (e.g. Taiwan and Singapore) tend

472   to exhibit an alveolar-retroflex phoneme contrast merger, or deretroflexion. Indeed, one recent study

473   documenting the alveolar-retroflex contrast in Singapore Mandarin has revealed that Singapore Mandarin

474   speakers show signs of deretroflexion, as evidenced by smaller acoustic differences between the two

475   categories of phonemes as compared to Beijing Mandarin speakers (Goh et al., 2022). Therefore, it is

476   possible that there may be less structural overlap between Mandarin and Hindi retroflexion than would

477   be necessary for a learning transfer effect to be observed in adults (e.g. Chládková & Šimáčková, 2021).

478   Taken together, this suggests that the linguistic feature canonically known as 'retroflexion' may be under-

479   specified for the purposes of fine-grained perception tasks involving speakers of different languages.

480   Furthermore, merely having a contrast between dental and another place-of-articulation based on tongue

481   position (bunched or curled) is not a sufficient source of linguistic structure to enhance learning for a

482   subtle new acoustic contrast in the general population.

483   While it seems a small number of individuals do show the expected effect (as in the pilot study), they

484   were sufficiently rare in the general population, that no significant interaction was observed between

485   learning and self-reported Mandarin understanding. Indeed, some studies have also observed no

486   relationship between Mandarin speakers' general ability to understand Mandarin and their perceptual

487   sensitivity to the Mandarin retroflex (Goh & Styles, 2022), nor their production of Mandarin retroflexion

488   (Chung, 2006; Goh et al., 2022), suggesting that measures of self-reported language understanding may

489   not be sensitive to individual differences in phoneme perception.

490   Our small-scale pilot study was used to plan a preregistered study with sample size determined through

491   a combination of minimum sample for statistical procedures (Exploratory Factor Analysis) and Bayesian

492   stop rule. The combination of these tools gives us greater confidence that although some individuals in

493   the study did show behaviours consistent with 'learning' of the novel contrast (as in the pilot), the effect

494   did not generalize to the broader population. This finding suggests that a language-specific advantage for

495   learning statistical structure is not sufficiently prevalent or powerful to induce learning of a novel contrast

496   in *all* bilingual adults when the to-be-learned contrast is acoustically subtle, and does not overlap

497   substantially with phonetic features in their repertoire.

498

499 **Conclusion**

500   Many streams of research suggest that bilinguals have certain cognitive advantages over monolinguals,

501 and evidence is particularly compelling that learning of novel linguistic structures is enhanced by overlap

502 between known languages and to-be-learned linguistic features. In our statistical learning paradigm, we

503 found evidence of neither a statistical learning effect nor language specific exposure effects on learning

504 in the general population, in our large-scale preregistered sample. However, we did observe that *some*

505 Singaporean adult participants exhibit sensitivity to the unfamiliar Hindi contrast following bimodal

506 training, as shown in the results of the small-scale pilot, but these individuals do not appear to be

507 representative of the population in general. Since sensitivity to unfamiliar speech sound contrasts is

508 known to decline with age, this paradigm likely has its maximum effect in a younger age group, such as

509 school aged children (Evans et al. 2009; Vandermosten et al. 2018), or in infants (Maye et al., 2002; Saffran

510 2020). Moreover, in order to investigate the possibility of a language-specific learning transfer effect in

511 Singapore English-Mandarin bilinguals, training stimuli should be more closely aligned to the articulatory

512 characteristics typical to Singapore Mandarin. Further investigation with Singaporean English-Mandarin

513 bilingual children and a different set of training stimuli would allow us to find out whether differences in

514 individual familiarity with phonetic features does indeed interact with ability to detect and learn linguistic

515 contrasts.

516   **References**

517 Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. C. M. (2015). The bilingual advantage in phonetic learning.

518   *Bilingualism: Language and Cognition, 18*(4), 683 – 695. doi:10.1017/S1366728914000777

519 Antoniou, M. (2019). The advantages of bilingualism debate. *Annual Review of Linguistics, 5*, 395 – 415.

520   https://doi.org/10.1146/annurev-linguistics-011718-011820

521   Atkinson, R. (Executive Producer), & Senior, K. (Executive Producer). (2003). Mr Bean: The animated series.

522   (Series 2). [TV series]. FremantleMedia Enterprises.

523   Bartolotti, J., Marian, V., Schroeder, S. R., & Shook, A. (2011). Bilingualism and inhibitory control influence

524   statistical   learning   of   novel   word   forms.   *Frontiers   in   Psychology,*   2.

525   https://doi.org/10.3389/fpsyg.2011.00324

526   Best, C.T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception

527   of approximants. *Journal of Phonetics, 20*, 305–330.

528   Bialystok, E., & Craik, F. I. M. (2010). Cognitive and linguistic processing in the bilingual mind. *Current*

529   *Directions in Psychological Science, 19*(1), 19 – 23. https://doi.org/10.1177/0963721409358571

530   Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children.

531   *Developmental Science, 11*(2), 282 – 298. doi: 10.1111/j.1467-7687.2008.00675.x

532   Chládková, K., & Šimáčková, Š. (2021). Distributional learning of speech sounds: An exploratory study into

533   the effects of prior language experience. *Language Learning, 71*(1), 131–161. doi: 10.1111/lang.12432

534   Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*,

535   781. doi: 10.3389/fpsyg.2014.00781

536   Dienes, Z., (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology, 72*,

537   78 – 89. doi: 10.1016/j.jmp.2015.10.003.

538   Dienes Bayes factor calculator. Retrieved from [https://bencepalfi.shinyapps.io/Dienes_BF_calculator/].

539   Enomoto, K. (1994). L2 perceptual acquisition: The effect of multilingual linguistic experience on the

540   perception of a 'less novel' contrast. *Edinburgh Working Papers in Applied Linguistics, 5*, 15–29.

541   Retrieved from https://eric.ed.gov/?id=ED373553enmoto

542   Escudero, P., Hillenbrand, J., Benders, T., Wanrooij, K., & Hillenbrand, J. (2011). Enhanced bimodal

543   distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of*

544   *America, 130*(4), EL206–EL212. https://doi.org/10.1121/1.3629144

545 Escudero, P., & Williams, D. (2014). Distributional learning has immediate and long-lasting effects.

546 *Cognition, 133*(2), 408–413. https://doi.org/10.1016/j.cognition.2014.07.002

547 Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language

548 impairment. *Journal of Speech, Language, and Hearing Research*, *52*(2). https://doi.org/10.1044/1092-

549 4388(2009/07-0189)

550 Goh, H. L., Styles, S. J., & Onnis, L. (2020, September 9). Effect of unattended distributional training on

551 phoneme category discrimination in English-Mandarin bilingual adult participants in Singapore.

552 osf.io/s6vdn. Embargoed until September 13, 2024.

553 Goh, H. L., Woon, F., Moisik, S. R., & Styles, S. J. (2022, May 10). Contrastive alveolar/retroflex phonemes

554 in Singapore Mandarin bilinguals: Comprehension rates for articulations in different accents, and

555 acoustic analysis of productions. PsyArXiv. https://doi.org/10.31234/osf.io/7szcu

556 Golestani, N., Paus, T., & Zatorre, R. J. (2002). Anatomical correlates of learning novel speech sounds.

557 *Neuron, 35*(5), 997 – 1010. doi: 10.1016/s0896-6273(02)00862-0

558 Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language

559 phonology. *Brain and Language, 109*(2–3), 55–67. doi: 10.1016/j.bandl.2008.01. 005.

560 Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes.

561 *Second Language Research, 23*(1), 65 – 94. doi: 10.1177/0267658307071601

562 Kaushanskaya, M., & Marian, V. (2009). Bilingualism reduces native-language interference during novel-

563 word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(3), 829–835.

564 https://doi.org/10.1037/a0015275

565 Kemp, C. (2007). Strategic processing in grammar learning: Do multilinguals use more strategies?

566 *International Journal of Multilingualism, 4*, 241–261. https://doi.org/10.2167/ijm 099.0

567   Kopečková, R. (2016). The bilingual advantage in L3 learning: a developmental study of   rhotic   sounds.

568   *International   Journal   of   Multilingualism,   13:4*,   410   –   425.

569   https://doi.org/10.1080/14790718.2016.1217605

570   Kovács, Á. M., & Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants.

571   *Science, 325* (5940), 611–612. https://doi.org/10.1126/science.1173947

572   Ladefoged, P., & Maddieson, I. (1996). *The sound of the world's languages*. Oxford: Blackwell Publishing.

573   Lakens, D., McLatchie, N., Isager, P.M., Scheel, A. M., & Dienes, Z. (2018). Improving inferences about null

574   effects with Bayes factors and equivalence tests. *The Journals of Gerontology Series B Psychological*

575   *Sciences and Social Sciences, 75*(1), 45 – 57. doi: 10.1093/geronb/gby065.

576   Linares, D., & López-Moliner, J. (2016). quickpsy: An R Package to Fit Psychometric Functions for Multiple

577   Groups. *The R Journal, 8*(1), 122–131. https://journal.rproject.orgarchive/2016-1/linares-na.pdf.

578   Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder

579   for the social sciences. *Behavior Research Methods, 44*(2), 314 - 324.

580   Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect

581   phonetic discrimination. *Cognition*, *82*(3), B101–B111. https://doi.org/10.1016/S0010-0277(01)00157-3

582   Oh, J. S., Au, T. K.-F., & Jun, S.-A. (2010). Early childhood language memory in the speech perception of

583   international   adoptees.   *Journal   of   Child   Language,   37*(5),   1123–1132.

584   https://doi.org/10.1017/S0305000909990286

585   Onnis, L., & Thiessen, E. (2013). Language experience changes subsequent learning. *Cognition, 126*(2), 268

586   – 284. doi: 10.1016/j.cognition.2012.10.008

587   Onnis, L., Chun, W. E., & Lou-Magnuson. (2018). Improved statistical learning abilities in adult bilinguals.

588   *Bilingualism: Language and Cognition, 21*(2), 427 – 433. doi: 10.1017/S1366728917000529

589   Ou, S., & Guo, Z. (2014). Mandarin retroflex sounds perceived by non-native speakers. *Chinese Collections*

590   *of   National   Taipei   University   of   Education,   42   –   76.*

591   *https://tpl.ncl.edu.tw/NclService/pdfdownload?filePath=lV8OirTfsslWcCxIpLbUfvnJVVyS2MdWoteG0es*

592   *3wArWL458qQ1UxEbw9UWP4zbv&imgType=Bn5sH4BGpJw=&key=9_Jt0kA3BKHdUYtXbhriYY2qqWybc*

593   *x2_q1C2oxESAeVVU9OyI NO4qBZJhLTxWd&xmlId=0006 794641*

594   R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical

595   Computing, Vienna, Austria. https://www.R-project.org/.

596   Saffran, J., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., & Cushman, F. (2008). Grammatical pattern

597   learning by human infants and cotton-top tamarin monkeys. *Cognition*, *107*(2), 479–500.

598   https://doi.org/10.1016/j.cognition.2007.10.010

599   Singh, L., Fu, C. S. L., Tay, Z. W., & Golinkoff, R. M. (2018). Novel word learning in bilingual and monolingual

600   infants: Evidence for a bilingual advantage. *Child Development, 89*(3), e183–e198.

601   https://doi.org/10.1111/cdev.12747

602   Singh, L., & Seet, S. K. (2019). The impact of foreign language caregiving on native language acquisition*.*

603   *Journal of Experimental Child Psychology, 185*, 51–70. https://doi.org/10.1016/j.jecp.2019.04.010

604   Tees, R. C., & Werker, J. F. (1985). Perceptual flexibility: Maintenance or recovery of the ability to

605   discriminate non-native speech sounds. *Canadian Journal of Psychology, 38*(4), 579 – 590. doi:

606   10.1037/h0080868

607   Terry, J., Ong, J. H., & Escudero, P. (2015). Passive distributional learning of non-native vowel contrasts

608   does not work for all listeners. *Proceedings of the 18th International Congress of Phonetic Sciences,*

609   *Glasgow, Scotland 2015.* http://www.internationalphoneticassociation.org/icphs-

610   proceedings/ICPhS2015/Papers/ICPHS0867.pdf

611   Tremblay, M. -C., & Sabourin, L. (2012). Comparing behavioural discrimination and learning abilities in

612   monolinguals, bilinguals and multilinguals. *The Journal of the Acoustical Society of America, 132*(5), 3465

613   – 3474. doi: 10.1121/1.4756955

614   Vandermosten, M., Wouter, J., Ghesquiere, P., & Golestani, N. (2018). Statistical learning of sounds in

615   dyslexic and typical reading children. *Scientific Studies of Reading, 23*(1), 1–12. doi:

616   10.1080/10888438.2018.1473404

617   Verma, R, & Chawla, P. (2003, May 9–11). Comparative analysis of Hindi retroflex and dental CV syllables

618   and their synthesis. [Paper presentation]. International Speech Communication Association, Mumbai,

619   India. Retrieved from https://www.iscaspeech.org/archive_open/wslp_03/w slp_1 63.pdf

620   Wang, T., & Saffran, J. R. (2014). Statistical learning of a tonal language: The influence of bilingualism and

621   previous linguistic experience. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00953

622   Wanrooij, K., Escudero, P., & Raijmakers, M. E. J. (2013). What do listeners learn from exposure to a vowel

623   distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics, 41*(5),

624   307–319. https://doi.org/10.1016/j.wocn.2013.03.005

625   Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual

626   reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49 – 63. doi:

627   10.1016/S0163-6383(84)80022-3

628   Werker, J. F. (1986). The effect of multilingualism on phonetic perceptual flexibility. *Applied

629   Psycholinguistics, 7*(2), 141–155. https://doi.org/10.1017/S0142716400007360

630   Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How do infants become experts at native-speech

631   perception? *Current Directions in Psychological Science*, *24*(4), 221 – 226. doi:

632   10.1177/0963721412449459

633   Yoshida, H., Tran, D. N., Benitez, V., & Kuwabara, M. (2011). Inhibition and adjective learning in bilingual

634   and monolingual children. *Frontiers in Psychology, 2*. https://doi.org/10.3389/fpsyg.2011.00210

Figure 1

Schematic of stimuli presented during a practice trial

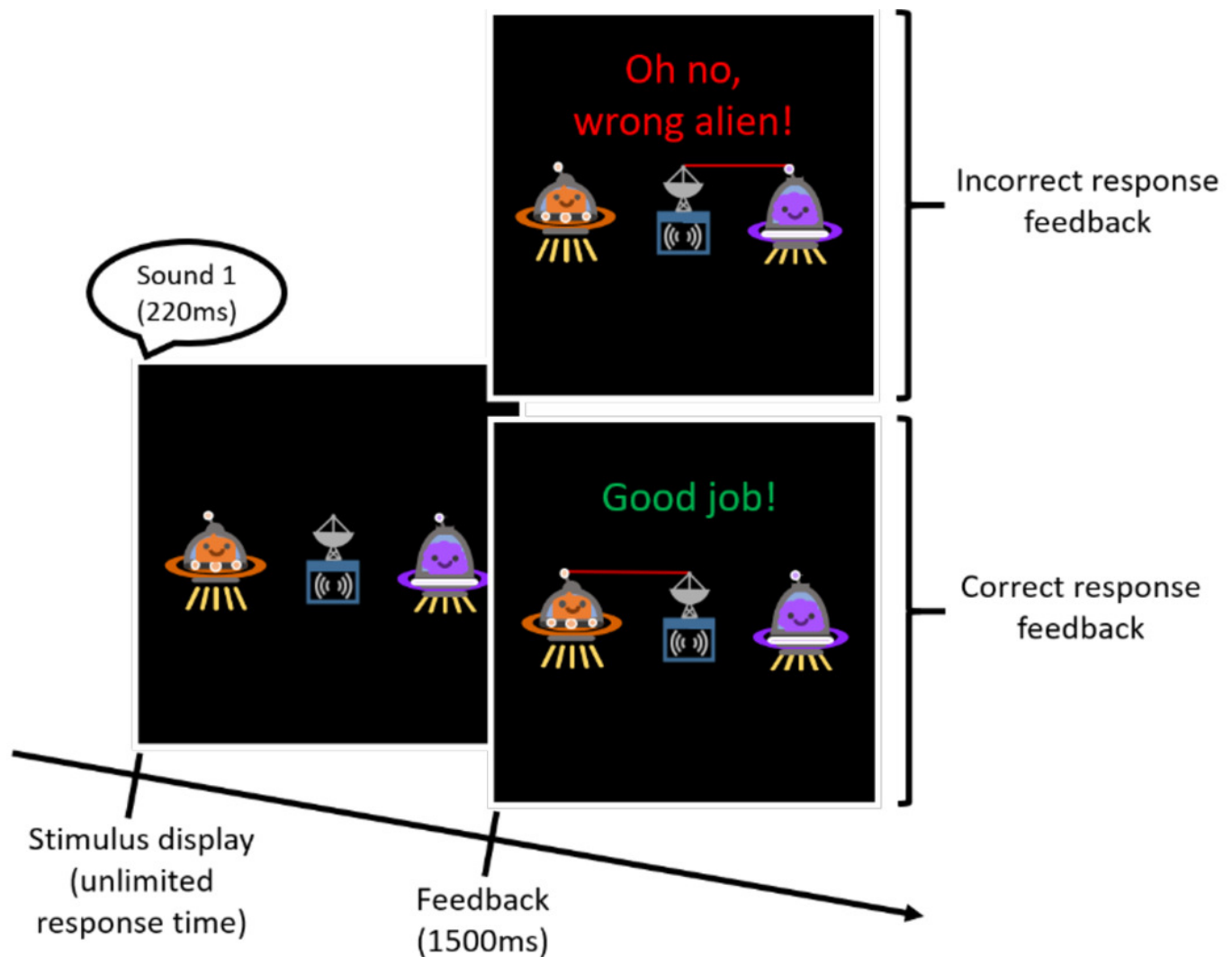



Figure 2

Frequency of sound tokens in bimodal distribution used for training.

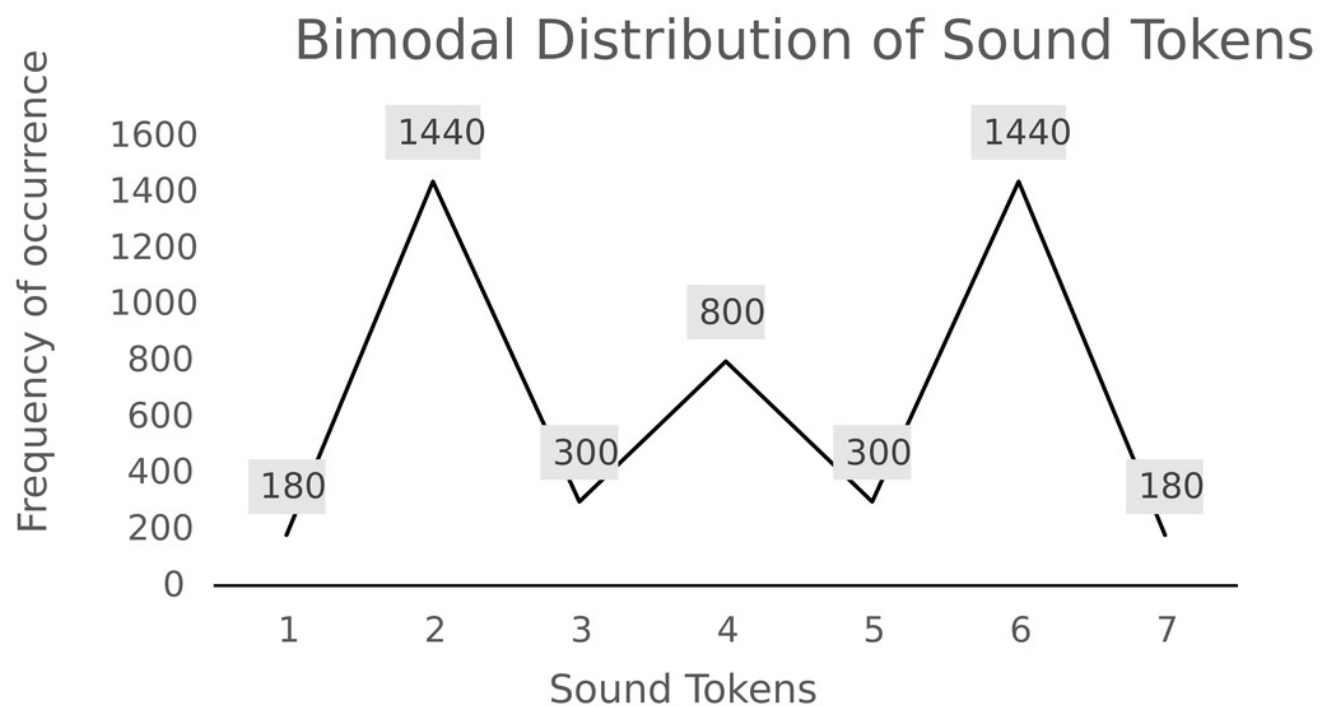



Figure 3

Self-Rated English and Mandarin Understanding Ability

Pilot participants' self-reported English and Mandarin understanding ability ($N = 15$).
Responses jittered by .25 for visualization.

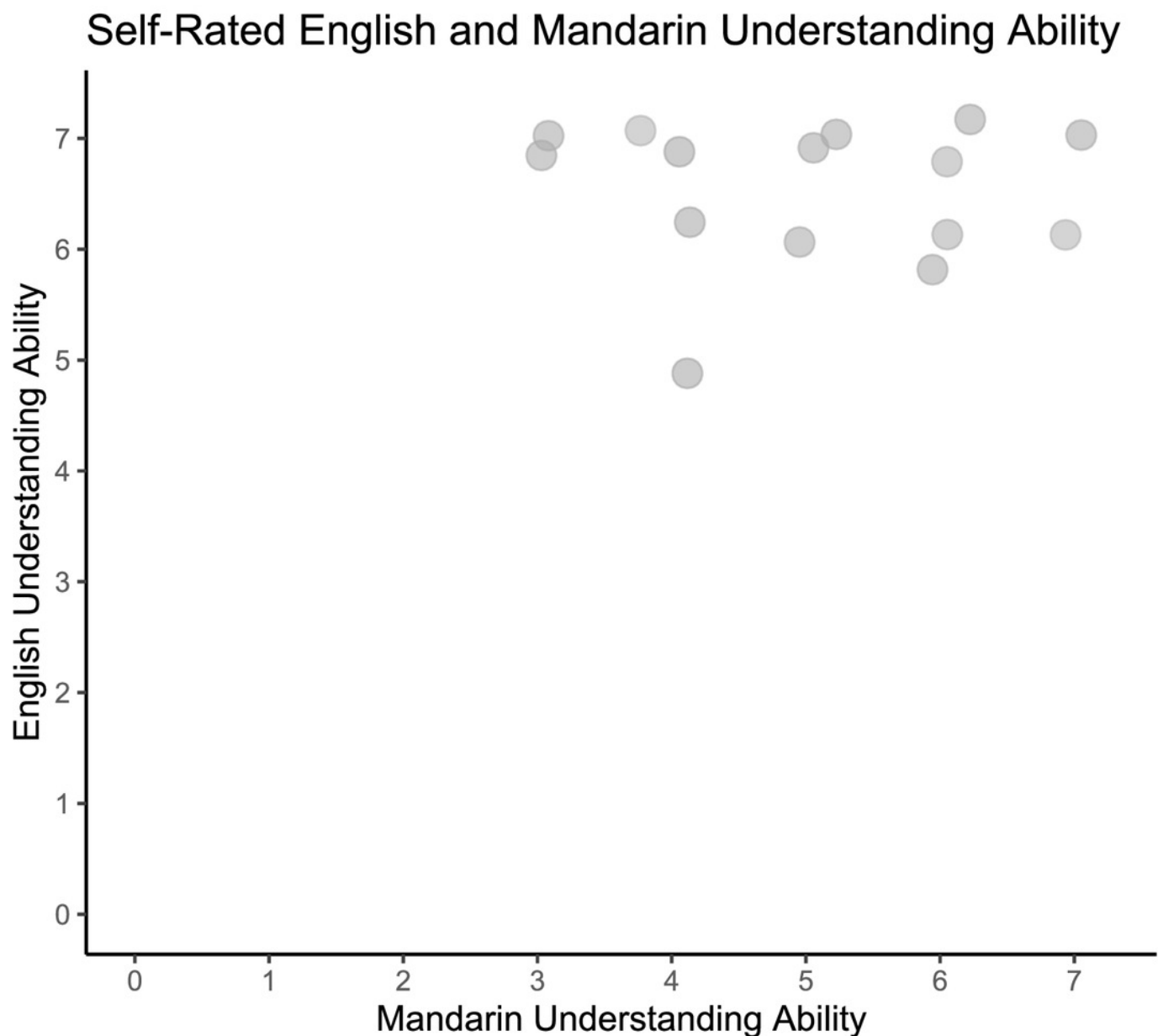



Figure 4

Categorisation sensitivity slopes for individual participants

Each slope shows the fitted psychometric function in a) pre-training test and b) post-training test (N = 15).

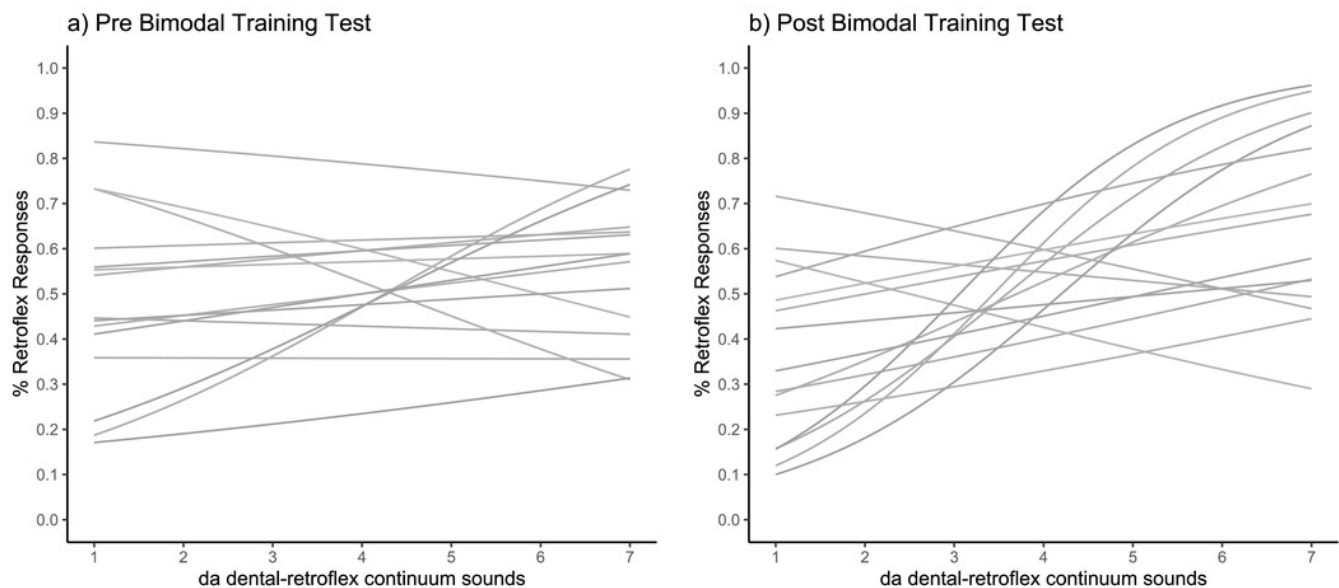



Figure 5

Self-Rated English and Mandarin Understanding Ability

Distribution of participants' English and Mandarin proficiencies in main study ($N = 50$).

Responses jittered by .25 for visualization.

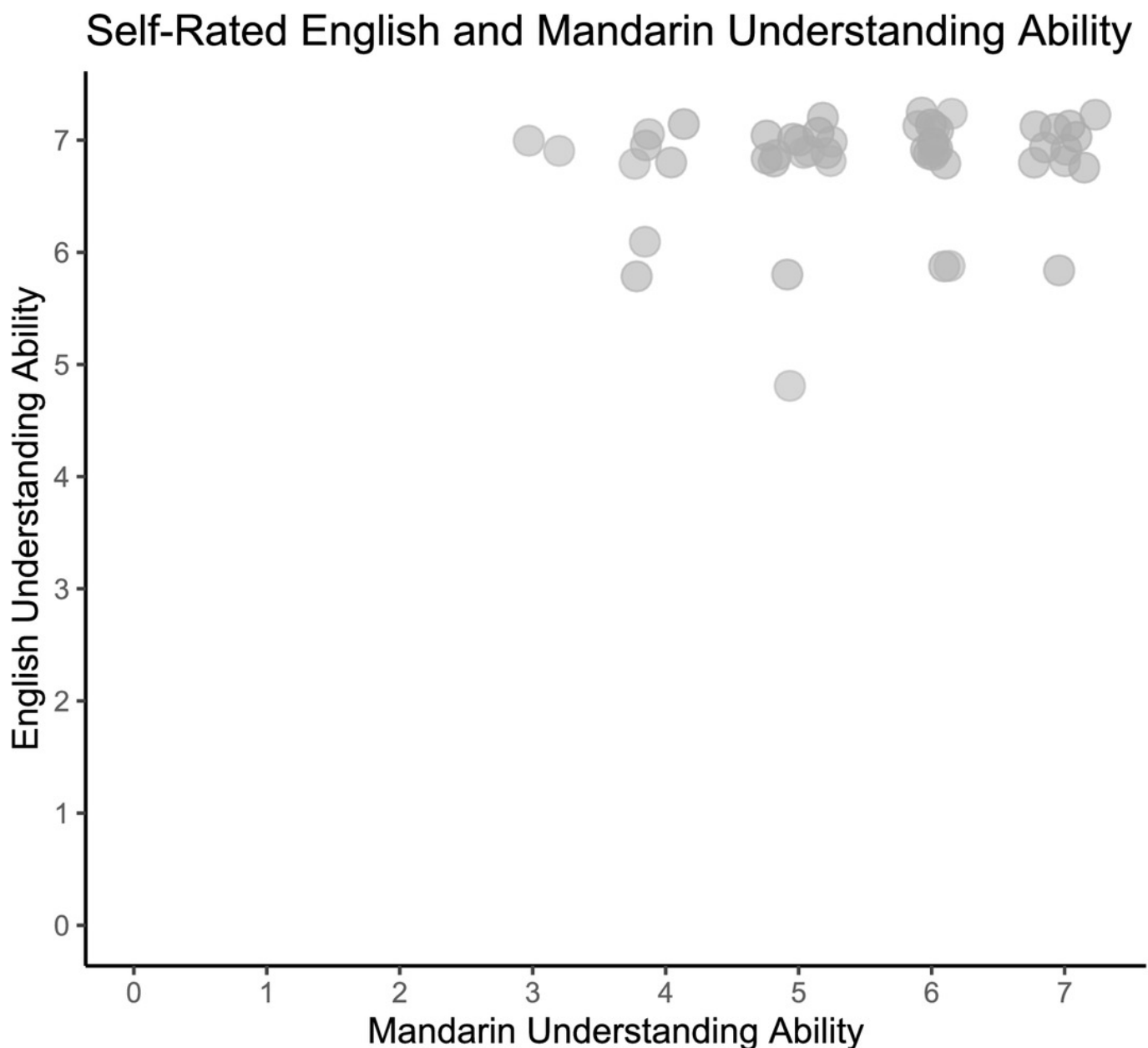



Figure 6

Categorisation sensitivity slopes for individual participants

Each slope shows the fitted psychometric functions for a) pre-training test and b) post-training test ($N = 50$)

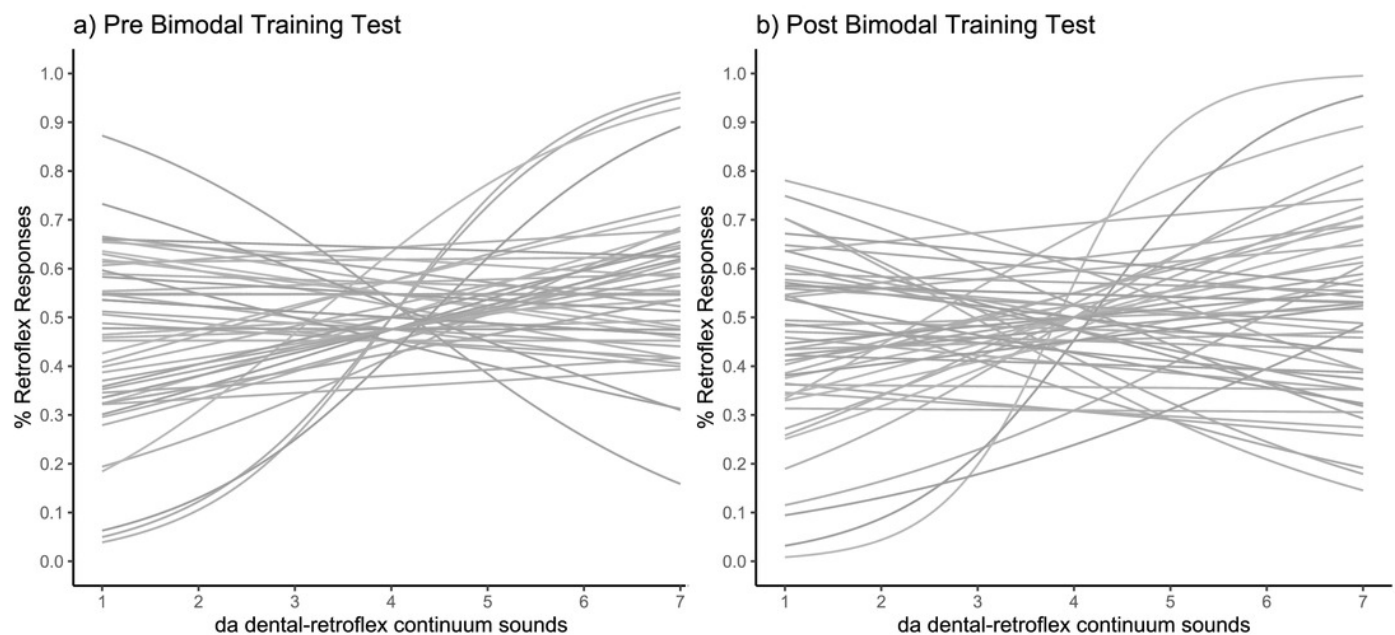



Figure 7

Interaction Between Learning Effect and Self-Rated Mandarin Understanding Ability

Fitted slope values for individuals at pre- and post- training, coloured by self-reported understanding ability of Mandarin Chinese (1 = 'a few words'; 7 = 'native').

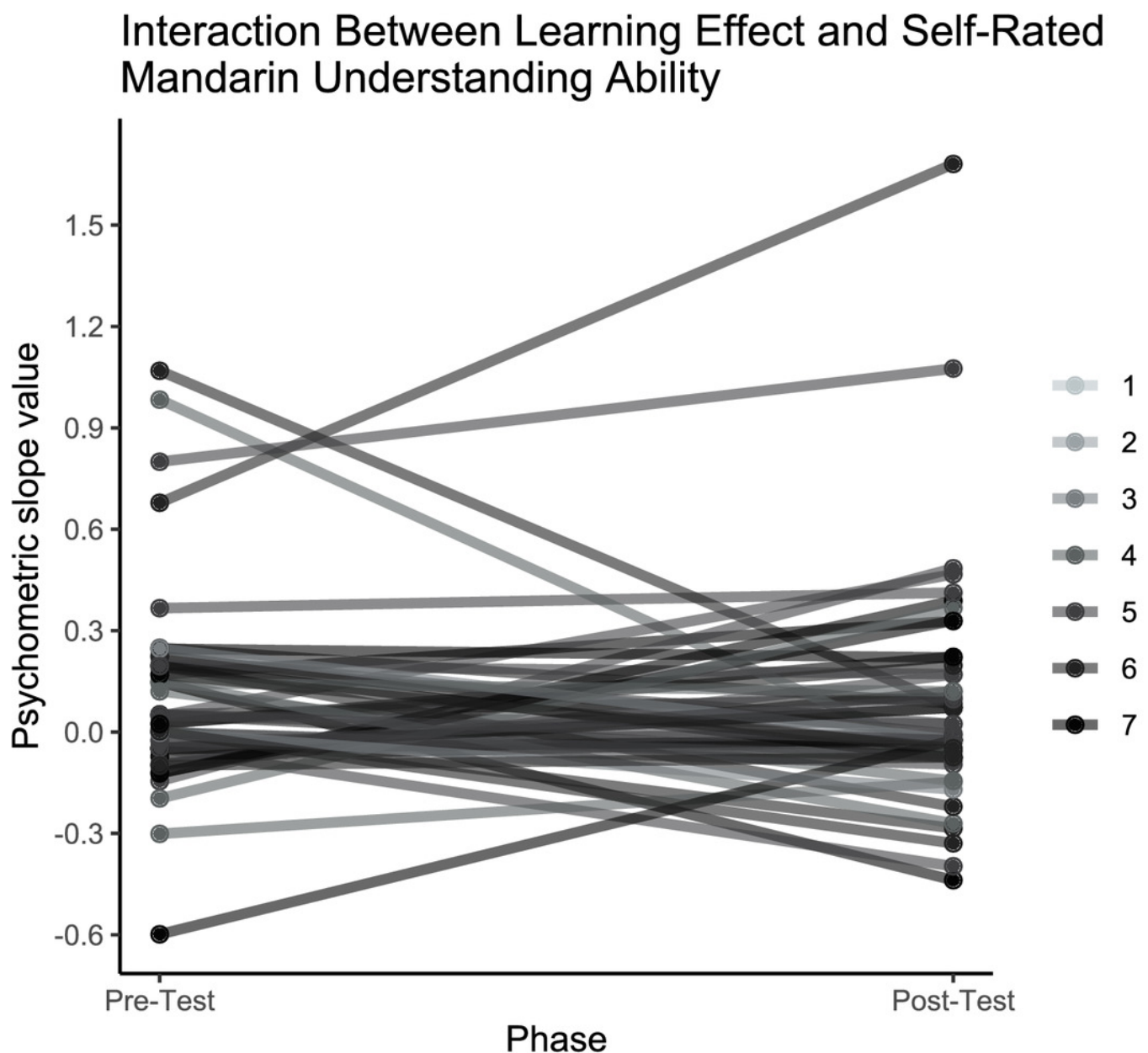
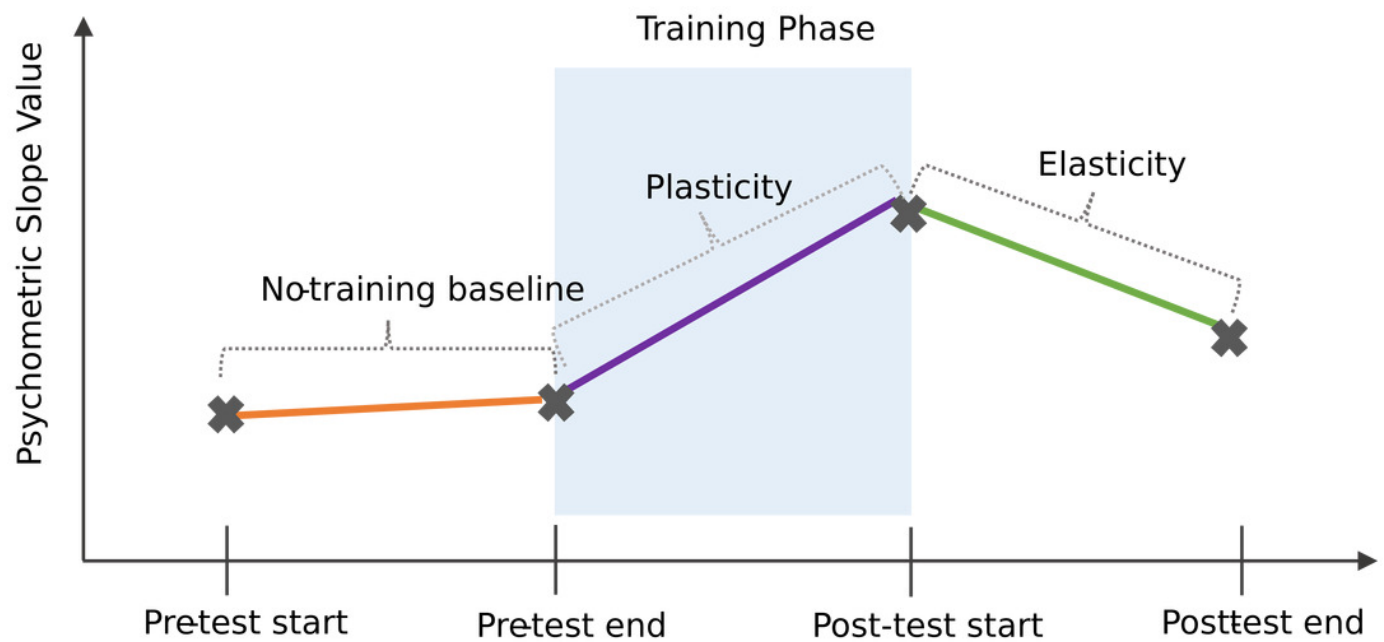



Figure 8

Schematic of a possible learning microstructure across distributional learning task.

Factor loadings on exploratory factor analysis of learning microstructure and language understanding.

1   Table 1: Factor loadings on exploratory factor analysis of learning microstructure and language

2   understanding.

	Component		
	1	2	3
Eigenvalues	2.41	1.19	.92
Percent Variance Explained	40.2%*	19.8%	15.4%
Factor loadings			
No-Training Baseline	.663		
Plasticity	-.474		
Elasticity		.687	.507
Mandarin Understanding	-.419	.681	-.562
No. Languages Understood	-.802		
Language Understanding score	-.914		

3   (Note: Elasticity is inversely represented, with positive value indicating lower Elasticity.

4   * Indicates principal components of interest that account for >20% of variance.)