NANYANG
TECHNOLOGICAL
UNIVERSITY
SINGAPORE

LANGUAGE-SPECIFIC DISTRIBUTIONAL LEARNING ADVANTAGES IN SINGAPORE ENGLISH-MANDARIN BILINGUAL ADULTS

HANNAH L. GOH

INTERDISCIPLINARY GRADUATE PROGRAMME

2023

LANGUAGE-SPECIFIC DISTRIBUTIONAL LEARNING ADVANTAGES IN SINGAPORE ENGLISH-MANDARIN BILINGUAL ADULTS

HANNAH L. GOH

INTERDISCIPLINARY GRADUATE PROGRAMME

A thesis submitted to the Nanyang Technological University in
partial fulfilment of requirement for the degree of Doctor of
Philosophy

2023

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

17 July 2023

. . . .
Date

. . . .
Hannah L. Goh

Supervisor Declaration Statement

I have reviewed the content of this thesis and to the best of my knowledge, it does not contain plagiarised materials. The presentation style is also consistent with what is expected of the degree awarded. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accordance with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

17 July 2023

. . . .

Date

. . . .

Assoc Prof Suzy Styles

Authorship Attribution Statement

This thesis contains material from 1 paper published in the following peer-reviewed journal – PeerJ – in which I am listed as an author. This thesis also contains material from 1 paper which is at the preprint stage and can be found on PsyArXiv – in which I am listed as an author.

The studies reported in Chapter 2 are also reported in Goh, H. L., Onnis, L., & Styles, S. J. (2023). 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. *PeerJ, 11*, e15467. https://doi.org/10.7717/peerj.15467.

The studies reported in Chapter 4 are also reported in Goh, H. L., Woon, F. T., Moisik, S. R., & Styles, S. J. (2022, May 10). Contrastive alveolar/retroflex phonemes in Singapore Mandarin bilinguals: Comprehension rates for articulations in different accents, and acoustic analysis of productions. https://doi.org/10.31234/osf.io/7szcu

Preliminary analyses of the studies in Chapter 2, Chapter 5, and Chapter 6 were reported at the following conferences: The Cognitive Science Society Conference (2021), The Cognitive Science Society Conference (2022), The Cognitive Science Society Conference (2023).

The contributions of the co-authors for the study in Chapter 2 are as follows:

- Prof Suzy J. Styles provided the initial project direction.
- I prepared the manuscript drafts for the published papers, in consultation with Prof Suzy J. Styles and revision by Prof Luca Onnis.
- I designed the study in collaboration with Prof Suzy J. Styles and Prof Luca Onnis.
- I conducted the data analyses with guidance from Prof Suzy J. Styles.
- All study data were collected by me.

The contributions of the co-authors for the study in Chapter 4 are as follows:

- I worked on the initial project direction with Prof Suzy J. Styles.
- I prepared the manuscript drafts in consultation with Prof Suzy J. Styles. The manuscript was revised by Fei Ting Woon, Prof Suzy J. Styles, and Prof Scott R. Moisik.
- I co-designed the study with Fei Ting Woon, Prof Suzy J. Styles, and Prof Scott R. Moisik.
- I conducted the data analyses with guidance from Prof Scott R. Moisik and Prof Suzy J. Styles.
- All study data were collected by me.

17 July 2023

. . . .

Date Hannah L. Goh

Acknowledgements

I am immensely grateful to my supervisor Prof Suzy Styles for her wonderful guidance. Her knowledge and expertise have played a huge role in helping me find my footing in the fields of cognitive psychology and psycholinguistics, and her passion for understanding the intricacies of under-documented languages is truly infectious. Without her dedicated mentorship, I would not have had such a fulfilling research experience. Thank you, Suzy!

I would also like to thank the members of my thesis advisory committee, Prof Annabel Chen and Prof Scott Moisik for their thoughtful feedback at each stage of this project – thank you for your valuable time and insight. Special thanks to Scott for showing me the linguistic ropes in acoustic analysis and stimulus creation, and letting me use your ultrasound equipment.

I will always be grateful to Dr. James Mantell and Dr. Gerrit Maus for first sparking my interest in cognition, and for their continued encouragement in my academic pursuits. I would also like to thank everyone who has shown me unconditional care and assistance – the amazing people in BLIP Lab, BCS, and all my friends and family, especially Deborah, Shaza, Esh, Vic, Anna, and Tea. Thanks also to American rock bands Thursday and My Chemical Romance for the emotional support and for reuniting when I needed them most.

Finally, I would like to thank my cat, Ollie, for his constant company through all the late nights of writing, and for waking me up bright and early every morning so I could get to work on this thesis. You’re a good cat, Ollie.

Table of Contents

List of Tables

List of Figures

Abstract

In this thesis, we set out to determine if we could find evidence for language learning advantages in bilingual adults that stem from language-specific transfer effects. The language-specific hypothesis of bilingual learning advantages suggests that language learning advantages are tied to an individual's linguistic experience. Thus, learning advantages are likely to stem from transfer effects between known languages, and to-be-learnt language stimuli. We focused our examination of this question on English-Mandarin bilingual adults in Singapore, as they consist of a largely homogenous group of bilinguals with intragroup differences in the language balances of their language repertoire (particularly for Mandarin Chinese). We chose to utilise a distributional learning paradigm in this investigation as such paradigms have been shown to be a robust method of assessing language-specific learning effects in adults (e.g., Chládková & Šimáčková, 2021; Chládková et al., 2022). Based on the language-specific hypothesis of bilingual learning advantages, we reasoned that individual variance in Mandarin proficiency would act as a predictive factor in determining individual differences in learning effects following bimodal distributional training on stimuli that share featural similarities with Mandarin Chinese. In essence, individuals with higher understanding abilities in Mandarin Chinese would demonstrate stronger learning effects for novel language stimuli that share phonemic features with Mandarin.

In this project, we conducted a total of six studies to thoroughly address this question. In Chapter 2 and Chapter 7, we directly examine this main research question. To fine-tune our investigation, we conducted four studies (Chapter 3, Chapter 4, Chapter 5, Chapter 6) documenting unique aspects of perception and production for the alveolar-retroflex contrast in the under-documented outer-circle variety of Singapore Mandarin. Overall, this project contributes novel insight into the validity of the language-specific learning hypothesis in bilinguals, and our documentations of Singapore Mandarin provide fresh details on the nature of the alveolar-retroflex contrast in Singapore Mandarin with the use of fine-grained acoustic analyses and behavioural measures.

Chapter 1 lays the basis for our main research question on the language-specific hypothesis of bilingual learning advantages. In Section 1.1, we conduct a thorough review of existing literature on three main overarching hypotheses on the nature of language learning advantages in bilinguals and multilinguals. We justify our interest in pursuing an investigation on the language-specific hypothesis. In Section 1.2, we discuss the unique linguistic landscape of Singapore, and detail how resident English-Mandarin bilinguals present us with a uniquely suitable group of individuals to examine this question with. Finally, in Section 1.3 we examine literature on distributional learning. We discuss studies on language acquisition via distributional learning in naturalistic settings, as well as in experimental settings and detail our reasoning for choosing a distributional learning paradigm in our investigation of the language-specific learning hypothesis of bilingual learning advantages.

Chapter 2 contains our first experimental exploration on the language-specific hypothesis of bilingual learning advantages via transfer effects. In this investigation, we trained English-Mandarin bilingual adults on a Hindi dental-retroflex phoneme contrast with a bimodal distributional learning paradigm. We reasoned that participants with higher Mandarin understanding proficiencies should show larger learning effects due to transfer effects as Hindi and Mandarin ostensibly both feature a dental/alveolar-retroflex place of articulation contrast (Hindi: stop consonants, Mandarin sibilant consonants). We examined overall learning effects in a pilot study (*N* = 15) and a main study (*N* = 50) and assess language-specific transfer effects in the main study. While we found evidence of overall learning in the pilot study, this was not replicable in our main study with a larger sample size, and we did not observe any language-specific transfer effects. A main confounding factor was identified from studies that were published after this study was completed. While formal Mandarin educational texts suggest that “retroflex” phonemes in Mandarin share a “curled” tongue place of articulation similar to that of Hindi retroflexion (Ou & Guo, 2014), recent lingual imaging studies show that Taiwan and standard Beijing Mandarin speakers are unlikely to produce “curled” retroflexion in speech at all (Luo, 2020; Tiede et al., 2019). We reasoned that this could have affected our participants’ ability to learn the Hindi contrast. This would also mean that the Hindi dental-retroflex stimulus set would be unsuitable for assessing a language-specific hypothesis of transfer effects in Singapore English-Mandarin bilingual adults.

To confirm if the “curled” retroflex is indeed absent in Singapore Mandarin, we then conducted an exploratory observational lingual ultrasound imaging study in Chapter 3 (*N* =5). The results of this study revealed that Singapore Mandarin speakers do not use a “curled” retroflex tongue position – their production of Mandarin retroflex phonemes, instead using a “humped” laminal post-alveolar tongue posture for Mandarin retroflexion. This study showed that we would need a more suitable set of stimuli to assess the language-specific hypothesis of bilingual learning advantages in Singapore Mandarin speakers. Studies (Chung, 2006; Chen et al., 2016) suggest that the phonological nature of the Singapore Mandarin alveolar-retroflex contrast likely differs from Mandarin varieties documented in existing Mandarin corpora (standard Mandarin: Sze et al., 2014; Hong Kong Mandarin: Tse et al., 2017) due to phoneme contrast merger (deretroflexion). In order to obtain a suitable stimulus set aligned to the unique norms of Singapore Mandarin, we thus conducted a detailed documentation of perception and production of the alveolar-retroflex contrast in contemporary Singapore Mandarin.

In Chapter 4, we discuss the details of this documentation conducted with novel speech elicitation and word identification tasks (*N* = 50). We examined the acoustic nature of the contemporary Singapore Mandarin alveolar-retroflex contrast with centre of gravity (CoG) analyses and general additive mixed modelling (GAMM) on long-term averaged spectra (LTAS) of for the first time in known literature. We also examined participants' word identification accuracy for alveolar-retroflex minimal pairs of words in Singapore and standard Beijing Mandarin. The results of the study revealed that contemporary Singapore Mandarin speakers produce alveolar and retroflex phonemes contrastively, albeit to a lesser extent than is typically reported for standard Beijing Mandarin (e.g., Chang, 2012). We also found that Singapore Mandarin speakers typically demonstrate very high levels of Mandarin word identification accuracy in both the local as well as the standard variety of Beijing Mandarin.

Using the acoustic information obtained in Chapter 4, we synthesised a novel 8-step alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast continuum based on the acoustic norms of contemporary Singapore Mandarin. We then conducted a study in Chapter 5 with a novel paradigm (the Spring-Village CROWN Game) to obtain fine-grained measures of perceptual sensitivity in the form of speech sound ambiguity resolution using our novel alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast continuum ($N = 62$). We also conducted a correlation analysis to determine if higher levels of perceptual contrast sensitivity are linked to higher Mandarin understanding abilities. The results of this study showed that the majority of contemporary Singapore Mandarin speakers perceive the alveolar and retroflex categories of speech sounds contrastively, albeit with a large range of individual differences in sensitivity for speech token ambiguity. We did not find evidence for a link between Mandarin understanding abilities and these individual differences in perception, suggesting that high levels of perceptual sensitivity for this phoneme contrast are not required for high levels of general speech comprehension in contemporary Singapore Mandarin.

We then conducted an exploratory study in Chapter 6 (*N* = 20) to examine links between perception and production of an alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast within our target population of Singapore English-Mandarin bilinguals. We used the speech elicitation task from Chapter 4 to collect measures of individual differences in production of the alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast, and the Spring-Village CROWN Game from Chapter 5 to measure individual differences in perception. The results of this study revealed a significant speech perception-production link tied to the retroflex affricate [ʈʂʰ] on the level of the individual. This link showed that individuals who produced their retroflex frication in a manner more similar to that of alveolar frication tended to have less sensitivity for retroflex cues in ambiguous speech sounds.

Finally, we returned to our main investigation of the language-specific hypothesis of bilingual learning advantages with our study in Chapter 7. We trained English-Mandarin bilingual adults on our novel synthesised alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast with a bimodal distributional learning paradigm. We reasoned that participants with higher Mandarin understanding proficiencies should show larger learning effects due to transfer effects between their real-world Mandarin experience and the acoustically complex training stimuli. We examined overall learning effects in a pilot study ($N = 20$) and a main study ($N = 50$) and assessed language-specific transfer effects in the main study. We found evidence of overall learning effects in both the pilot and the main study, confirming that adults *can* learn from unattended distributional training if there is sufficient overlap between their languages and the to-be-learnt stimuli. We also found evidence of a transfer effect tied to Mandarin understanding proficiency with larger learning effects being significantly associated with higher Mandarin understanding proficiencies. This is the first time evidence has been found to show that language-specific learning transfer effects can be strong enough to be observed on the level of the individual within the linguistic variations of a single group of largely homogenous bilingual adults.

Chapter 1: Literature Review

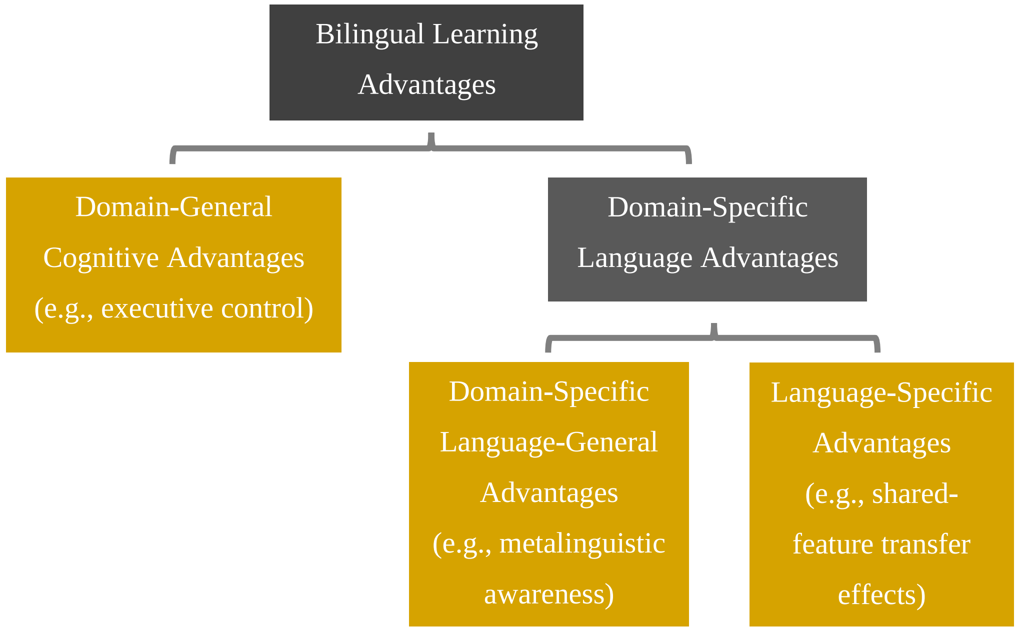
The ability to perceive and acquire new speech sounds is something which certain groups of people seem to be particularly adept at. Among these groups of people are bilinguals and multilinguals, who often seem to show learning advantages for acquiring perceptual sensitivity to novel speech sounds. We set out to investigate the underlying mechanisms that account for these bilingual learning advantages. Singapore is largely populated by bilinguals and multilinguals, and presents us with a unique context in which to conduct this investigation with the use of a distributional learning paradigm.

1.1 Bilingual Learning Advantage

Evidence in existing literature suggests that individuals with more than one language in their language repertoire appear to have certain learning advantages over monolinguals when it comes to the acquisition of novel language structures. Studies have shown that these advantages can be seen in tasks involving the learning of novel grammatical structures (e.g., Kemp, 2007; Kovács & Mehler, 2009; Kuo & Anderson, 2012; Sanz, 2000), the acquisition of novel vocabulary (e.g., Kaushanskaya & Marian 2009; Yoshida et al., 2011), and in developing sensitivities to novel language phoneme contrasts (Singh et al., 2018; Tremblay & Sabourin, 2012; Wang & Saffran, 2014). From an examination of current literature, we have identified three overarching hypotheses of interest that appear to account for different aspects of bilingual and multilingual language learning advantages. These accounts consist of domain-general cognitive advantages, language-general learning advantages, and language-specific learning advantages (as shown in Figure 1.1). In this chapter, we will examine these three bilingual learning advantage hypotheses and review research in support of each hypothesis.

Figure 1.1

Accounts for bilingual learning advantages identified in existing literature.



Note. Main three accounts of interest presented in yellow.

1.1.1 Domain-general accounts

Accounts of domain-general cognitive advantages typically propose that acquiring and using multiple languages inherently recruits certain cognitive functions for the purposes of executive control. Executive control plays a key role in language processing and production in bilinguals. Research suggests that all languages in a repertoire are simultaneously activated during language processing (Brysbalrt, 1998), and cognitive control over each specific language is needed to manage and inhibit which language is selected for speech processing and understanding as the situation demands. Evidence for the need for cognitive control in language processing in bilinguals can be seen in a study by Spivey and Marian (1999) which examined distraction caused by shared-language features in Russian and English. In their study, Spivey and Marian (1999) instructed Russian-English bilingual participants in Russian to move an object, while ignoring other objects on the table. They found that their bilingual participants typically showed a brief period of distraction due to the presence of task-irrelevant objects with English names that shared an initial sound with the name of the target object in Russian: e.g., being distracted by a marker while being instructed to move a stamp (*marku)* in Russian (Spivey & Marian, 1999).

At the same time, Emmorey et al. (2008) also demonstrated that this simultaneous activation of multiple languages occurs until late in the speech production stage where lexical inhibition is then introduced, by examining hearing American Sign Language (ASL)-English bilingual adults who grew up within deaf communities. In this study, the authors examined spontaneous language productions of their participants in conversation to determine if participants spent more time code-switching between their two languages or ‘code-blending’ both languages (i.e., simultaneously producing semantically equivalent ASL signs and spoken words). The authors proposed that if a dual activation of languages occurs until late in the speech production stage, and if lexical suppression is more cognitively costly as compared to activation, more instances of code-blending should be observed as compared to code-switching.

The results of their study indicated that this was indeed the case, with their participants frequently engaging in code-blending of semantically similar sign and spoken words, while rarely showing any instances of code-switching in their conversations.

Such studies on language activation models suggest that bilinguals and multilinguals have to manage and suppress shared structures between languages to successfully perceive and produce speech in each of their languages as the demands of the situation require. This is likely to be cognitively costly, and it has been suggested by researchers that this extensive amount of practice in cognitive control may lead to the development of a corresponding overall enhancement in domain-general cognitive functions (e.g., Bialystok, 2001). This may then in turn facilitate the acquisition and consolidation of new language structures into the existing language repertoire.

Indeed, some studies appear to demonstrate that bilinguals do display observable cognitive enhancements over monolinguals. For instance, Bialystok and colleagues conducted a series of studies which showed that bilingual children and adults typically have better executive control as compared to monolinguals, particularly in tasks involving cognitive conflict resolution. These include tasks such as the Dimensional-Change Card-Sort Task (Bialystok & Martin, 2004), a task involving interpreting ambiguity embedded in conflicting figures (Bialystok & Shapero, 2005), and the Stroop task (Bialystok et al., 2008), all of which require participants to inhibit conflicting perceptual information in order to successfully complete a task. At the same time, the notion that extensive experience with bilingualism leads to enhanced cognitive functions can also be seen in studies that compare different groups of bilinguals such as bilinguals who acquired both languages earlier in life as compared to later bilinguals. A study by Carlson and Meltzoff (2008) demonstrated that advantages in executive control are closely tied to the degree of experience that one has with being bilingual, with children who had been Spanish-English bilingual from infancy showing significantly higher levels of cognitive control on tasks of perceptual conflict resolution, as compared to later bilingual English-speaking children who had received six months of language immersion programmes in either Spanish or Japanese. From these studies, we can see that bilinguals do indeed appear to have certain enhancements in cognitive functioning. However, the results of these studies are largely based on measures of cognition that assess non-verbal, non-linguistic skills. As such, it is unclear how exactly domain-general accounts contribute to language-specific learning advantages often seen in bilinguals.

Studies have attempted to bridge this gap by examining correlations between tasks of language acquisition, and measures of executive control. For instance, Yoshida et al. (2011) examined the performance of 3-year-old bilinguals and monolinguals on a novel adjective-learning task and a non-linguistic measure of attentional control. They found that the bilingual children performed significantly better than the monolingual children at both tasks, suggesting that domain-general cognitive advantages found in bilinguals could indeed be involved in the facilitation of linguistic task performance. However, mixed evidence for a correlation between executive functions and language learning has been found in other studies. For instance, Bartolotti et al. (2011) examined novel language learning (with a Morse code-based distributional learning task) and inhibitory control (with a Stroop task) in adults with varying levels of bilingual experience. By manipulating the levels of conflicting cues present in the learning task, they found that bilingual experience was only related to enhanced learning performance under conditions of low cue conflict, while inhibitory control played a more crucial role regardless of bilingual experience in learning performance under conditions of highly conflicting cues. Thus, while the bilingual experience may be associated with cognitive benefits under some conditions, the presence of a mediating link between bilingualism and language learning nevertheless remains unclear.

At the same time, there remains an element of disagreement within current literature as to whether bilingualism is in fact associated with any cognitive advantages at all. A recent review of literature by Antoniou (2019) found that existing studies in the body of bilingual literature often conflict with each other on whether bilinguals do in fact show evidence of enhanced executive functions. In instances where there such evidence is found, it appears to disproportionately be seen in young children and the elderly, but less so in young adults (Antoniou, 2019). Moreover, studies on the link between bilingual experience and enhanced cognition do not always demonstrate a linear relationship between the two. For instance, a study by Yang (2017) examined

Korean-English bilinguals who either had intermediate or high proficiency in English and found, unexpectedly, that the intermediate proficiency bilinguals outperformed the high proficiency bilinguals at a test of working memory. Moreover, a large-scale study of executive functions in adults conducted by Paap et al. (2014) further shows that bilingualism and multilingualism may not be related to enhancements in various areas of executive functioning including inhibitory control and task conflict resolution. Therefore, it is difficult to determine the exact degree to which domain-general cognitive accounts might contribute to the language learning advantages observed in bilinguals.

1.1.2 Domain-specific accounts

Unlike domain-general accounts of bilingual language learning advantages, domain-*specific* accounts propose that bilingual learning advantages are inherently rooted in the linguistic experience of acquiring and managing multiple known languages. Thus, the cognitive benefits that arise from bilingualism, and which result in further language learning advantages, are specifically based in the domain of metalinguistic awareness. Crucially, enhanced metalinguistic awareness can result in an improvement in the ability to flexibly access and manipulate the underlying structural regularities of different languages – this plays a key role in facilitating the acquisition of novel language structures (Bialystok & Craik, 2010; Kuo & Anderson, 2008).

Support for enhanced bilingual metalinguistic awareness can be found in studies where bilinguals appear to show enhanced performance at extracting regularities in novel languages, and at manipulating this knowledge for application to new language stimuli, in studies of both children and adults. One example of this can be seen in a study by Kuo and Anderson (2012) who trained monolingual Mandarin-only, and bilingual Mandarin-Southern Min Taiwanese children on artificial two-syllable words based on a Mandarin-like phonological structure. In a 2-alternative-forced-choice test, bilingual children were significantly more accurate than the monolingual children at learning to identify legal and illegal words based on the phonotactic regularities that they were trained on. Furthermore, the bilingual children also displayed an ability to manipulate this information for application to novel stimuli, showing significantly better performance than the monolingual children at transferring these regularities to novel artificial words that they had not been trained on, but which shared similar phonotactic regularities.

Likewise, similar results showing evidence for domain-specific, language-general accounts have been found in studies of bilingual adults. For instance, Tremblay and Sabourin (2012) trained four groups of Canadian adult participants (monolingual, bilingual, multilingual, and untrained control) on a *voiced aspirated* Hindi dental-retroflex contrast, and tested the participants on their ability to discriminate tokens drawn from the contrast before and after training. The results of their study revealed that both the bilingual and multilingual groups displayed significantly better contrast discrimination as compared to the monolingual and control group participants, indicating that the experience of managing multiple languages in a repertoire was linked to the ability to successfully extract structural regularities from this novel language contrast. Tremblay and Sabourin (2012) then tested the same group of participants on their discrimination sensitivity for a novel untrained *voiceless unaspirated* Hindi dental-retroflex contrast, and found that bilingual participants outperformed the monolingual group, demonstrating evidence of an enhanced to ability to manipulate and transfer their structural knowledge of a known contrast to novel language stimuli that shared a similar place of articulation contrast but differed in the manner of articulation. This kind of transfer of linguistic knowledge from one type of linguistic structure to another provides a possible mechanism for language learning advantages in bilinguals. Evidence of enhancements in domain general language learning advantages can also be found in a study by Enomoto (1994) who tested monolingual and multilingual adult Japanese learners on their sensitivity to a perceptually difficult Japanese geminate contrast, and found that learners with a larger language repertoire displayed stronger sensitivity to this contrast as compared to English-speaking monolinguals. In particular, Enomoto (1994) found that this effect was larger for participants whose languages had structural overlaps with the Japanese geminate contrast in question, again suggesting that there might be transfer of knowledge from linguistic structures in one language to a novel learning context.

1.1.3 Language-specific accounts

Therefore, while there is some evidence for a domain-specific but *language-general* linguistic account of bilingual learning advantages rooted in metalinguistic awareness, it is likely that certain learning advantages might be *specific* to bilinguals' linguistic experience. As such, some learning advantages observed in bilinguals are likely to be tied to the specific featural overlaps between languages in their repertoire, and to-be-learnt language stimuli. In essence, bilinguals may have metalinguistic advantages that are specifically tuned to identifying familiar features in novel language stimuli. This could then feed into shared-feature transfer advantages in which bilinguals and multilinguals are able to transfer knowledge about shared language features (e.g., place of articulation, tone, and phonological structure) from known languages to novel language stimuli, thus facilitating learning effects (e.g., Kuo et al., 2016).

Evidence for a language-specific transfer effect in learning for bilinguals can be observed in studies that compare groups of participants with different language backgrounds. For instance, Bialystok et al. (2003) trained monolingual, Spanish-English bilingual, and Mandarin-English bilingual children on a phoneme segmentation task based on stimuli containing phonemic and morphological features that are unique to Spanish, and are not present English or Mandarin. They found that the Spanish-English bilingual children were significantly better at this task as compared to not only the English monolingual, but also the Mandarin-English bilingual children, thus demonstrating that the learning effects in the study were closely tied to linguistic experience, rather than to a language-general learning advantage for both groups of bilingual children.

At the same time, a study by Kuo et al. (2016) tested English monolingual, English-Japanese bilingual, and Japanese-English bilingual elementary school-aged children on their phonological awareness of real and artificial English-based words containing voiced onset phoneme contrasts that were either unique to English, or were shared between English and Japanese. In this study, both groups of Japanese-speaking bilingual participants outperformed the English monolinguals at tests of phonological awareness for artificial words that contained both Japanese and English voiced onset phonological segments. The Japanese-English bilinguals also had better phonological awareness for these stimuli as compared to the English-Japanese bilinguals. This shows that the transfer effects were stronger for both groups of bilinguals as compared to the English-speaking monolinguals, and that extensive experience with Japanese afforded some participants an additional language transfer advantage for the novel language stimuli. As such, we can see that the learning effects in this study were not only closely tied to the language repertoires of the participants, but were also enhanced by higher levels of relevant linguistic experience.

Similarly, Antoniou et al. (2015) trained and tested midwestern American English-speaking monolinguals, Mandarin-English bilinguals, and Korean-English bilingual adults on their perceptual sensitivity for a Mandarin-based retroflex phoneme contrast, and a Korean-based interdental fricative lenition phoneme contrast. The results of the study showed that only the Korean-English bilinguals were able to show learning effects for the Korean-based contrast, demonstrating clear evidence of a language-specific transfer effect in play. At the same time, however, the study also found that both the Mandarin-English and Korean-English bilinguals were able to successfully learn the Mandarin-based contrast. While this could be evidence for a language-general account of bilingual learning advantages, it is worth noting that the feature of retroflexion exists in Korean (e.g., in the retroflex variant of the Korean liquid /l/, Crosby & Dalola, 2021), and it is thus possible that the Korean-English bilinguals in Antoniou et al. (2015) were able to transfer knowledge of this feature to the novel Mandarin-based retroflex contrast.

Evidence of the role of language-specific transfer effects in bilinguals can also be found in situations of real-world language acquisition, such as in the case of Japanese learners in Enomoto’s (1994) study as discussed in the previous section. Additionally, evidence for language-specific transfer effects in naturalistic settings of language acquisition can also be found in a longitudinal three-year study conducted by Kopečková (2016) on different groups of bilinguals learning Spanish. This study found that bilingual Spanish learners who were already familiar with the alveolar approximant /ɹ/ phoneme in their language repertoire (e.g., German-Croatian, and German-Russian bilinguals) showed significantly more improvements in their production of the Spanish alveolar trill /r/ over time, as compared to bilinguals whose existing languages did not have an alveolar /ɹ/ (e.g., German-Israeli Hebrew speakers). This indicates that language-specific transfer advantages may extend to language acquisition in speech production for adults.

Besides this, some evidence suggests that language-specific transfer effects may be robust, lasting well into adulthood even if exposure to relevant languages only occurred early in life. An example of this can be seen in a study by Werker (1986) who examined the effect of early language exposure on phoneme contrast sensitivity in two groups of adults – one exposed to Hindi up to the age of two, and one with no exposure to Hindi across the lifetime. The two groups of participants were tested on their accuracy at discriminating a Hindi dental-retroflex contrast, and the results of the study showed that early exposure to Hindi was linked to significantly higher levels of discrimination accuracy for the contrast. Similarly, Oh et al. (2010) examined the effects of early language exposure in Korean-American adoptees who were only exposed to Korean in early childhood. Oh et al. (2010) found that individuals with early childhood exposure to Korean had a significant advantage at learning to identify Korean phonemes with aspiration and lenition as compared to individuals with no Korean exposure across the lifetime. Finally, evidence for the robustness of language-specific transfer effects following limited early language exposure can also be seen in the local context of adults in Singapore. In a recent study, Singh and Seet (2019) trained two groups of Singapore adults on Hokkien tonal contrasts. One group were raised by Hokkien-speaking caregivers up to the age of three, after which they had limited exposure to Hokkien, and the other group had no direct exposure to Hokkien during early childhood, and limited exposure over their lifetime. The results of their study revealed that adults who had been exposed to Hokkien early in life were significantly better at learning Hokkien tonal contrasts as compared to adults who had had no exposure to Hokkien in early life. Thus, we can see that even a limited amount of exposure to language features early in life may result in observable language-specific transfer effects that last well into adulthood.

Overall, the studies discussed in this section demonstrate that language learning effects in bilinguals often show evidence of being linked to shared linguistic features between known languages and to-be-learnt language stimuli, rather than being linked to domain-general advantages from the experience of simply being bilingual. Some of the studies discussed even show that the impact of transient language exposure early in life can still be seen in language-specific transfer effects later in adulthood. We therefore decided to investigate if evidence of such language-specific transfer effects can be found within the local population of bilinguals in Singapore.

The studies discussed in this chapter have largely examined language-specific learning effects by comparing different groups of bilinguals. While this has led to pertinent evidence for this account of bilingual learning advantages in the form of between-group comparisons, we believe that this methodology may overlook a rich source of data in bilingualism research by masking the individual linguistic variations found within ‘homogenous’ groups of bilinguals. This has implications on our current understanding of learning advantages in bilingualism, and limits the generalisability of these findings for cultural contexts where linguistic landscapes may be more nuanced, and may not fit the conventional view of what “typical” language backgrounds look like in conventional western research settings. We thus decided to instead examine bilingualism on its own as a continuous factor, and investigate if such effects might be observable within a single, largely homogenous group of bilinguals with varying levels of proficiency in their known languages. Due to the unique linguistic landscape of Singapore, the resident population of English-Mandarin bilinguals in Singapore presents us with a suitable group of bilinguals within which to investigate this question.

1.2 Bilingualism and Multilingualism in Singapore

1.2.1 Language mixes in Singapore

Singapore is a multi-ethnic city state in the Malay Peninsula of Southeast Asia with a resident population of 4.02 million people, and an ethnic composition of 74.1% Chinese, 13.6% Malay, 9.0% Indian, and 3.3% “Others” (i.e., Eurasian and other smaller minority groups; Cavallaro & Ng, 2014; DOS, 2022). The current linguistic landscape of the country comprises a large number of ‘English-knowing’ bilinguals whose main language repertoires consist of English and one of the officially designated ‘Mother Tongue’ languages (i.e., Mandarin Chinese for individuals of Chinese heritage, Bahasa Melayu for individuals of Malay heritage, and Tamil for individuals of Dravidian Indian heritage) (Cavallaro & Ng, 2014; Pakir, 1993; 2008). This particular pattern of bilingualism arose from a series of language policies introduced since the 1950’s to encourage the use of English as the country’s official language of governance and education, while also instituting a homogenous model of bilingualism in which English is treated as a primary language and an ethnically linked ‘Mother Tongue’ language is also supported (e.g., the Official Languages and National Language policy in the 1950’s, the Bilingualism policy in 1966; NLB, 2016). These policies have led to a drastic change in the linguistic landscape of the country within a few decades, such that large differences in language mixes can still be observed between generations of residents in the country.

Prior to the institution of these language policies, the linguistic landscape of the country before the 1980’s was highly varied, with residents of the country often having a “polyglossic” language repertoire of up to eight different languages (Cavallaro & Ng, 2014; Platt, 1980). In the early history of Singapore, two of the most commonly spoken home languages were Malay and Hokkien (a Southern Min Chinese language), with a pidgin variety of Malay – Bazaar Malay – used for cross-cultural communication between the different ethnic groups (Cavallaro & Ng, 2014; Gupta, 1998). English was rarely spoken by residents of the country, with only 1.8% of the population reporting being able to speak English in 1957 (Lee & Phua, 2020). There was also a considerable variety of commonly used vernacular languages. For instance, the majority of individuals of Chinese heritage spoke a range of South-Eastern and Southern Min Chinese languages such as Hokkien, Teochew, Hakka and Cantonese, with Mandarin Chinese only being spoken by 0.1% of the population in 1957 (Lee & Phua, 2020). Individuals of Malay heritage were also likely to speak a wide variety of home languages other than Bahasa

Melayu, including Boyanese, Javanese, Orang Seletar, and Baba (Peranakan) Malay (Cavallaro & Ng, 2014). This has since changed in the present day of Singapore, with remaining speakers of other vernacular languages often being from older generations of residents (DOS, 2021).

Among the language policies implemented in the country, education policies have played a particularly crucial role in modifying the languages that individuals are exposed to from a young age. For instance, English has been instituted as the mandatory language of instruction for all subjects in public schools since 1983 (NLB, 2016) following an increase in enrolment in English-medium schools beginning in 1965 (Dixon, 2005). As a result of this, proficiency in English is placed at a higher level of importance as compared to other locally spoken languages, as fluency in English is necessary for almost all areas of academic performance (Curdt-Christiansen & Sun, 2016). At the same time, language classes in the official 'Mother Tongue' languages have been part of public school curriculums since the Bilingualism policy of 1966, and were officially made a mandatory part of all forms of public schooling in 1987. In the case of Chinese languages, measures have also been taken to heavily discourage the use of Chinese home vernacular languages under the belief that exposing students to vernacular languages would make it harder for them to gain proficiency in the official 'Mother Tongue' language of Mandarin (Cavallaro & Ng, 2014). This belief was likely bolstered by an early study by the Ministry of Education which found that ethnically Chinese children who spoke primarily Chinese vernacular languages at home were less likely to perform well at their

‘Mother Tongue’ Mandarin examinations at school (Low, 1979, as cited in Teo, 2005). Among some of the measures taken to replace the use of vernacular Chinese languages with Mandarin include the banning of the use of ‘dialects’ on national television broadcasts in 1982 (Tan & Goh, 2011), and the introduction of the Speak Mandarin campaign in 1979.

The Speak Mandarin Campaign was first introduced in 1979, and utilised messages broadcasted from the government through the media to encourage ethnically Chinese Singaporeans to speak Mandarin instead of other vernacular home languages. Despite Hokkien originally being the most widely spoken Chinese language in the early history of the country, Mandarin was chosen as a new artificial ‘heritage’ lingua franca for the Chinese population of Singapore as it was perceived to be a more ‘prestigious’ language of culture in comparison to vernacular Chinese home languages which were framed by the ruling party as low-brow and vulgar (Lim et al., 2021). In particular, the variety of Hokkien spoken in Singapore was frowned upon for not being a “proper language” due in part to the fact that it had begun incorporating loan words from other languages into its vocabulary (Lock, 1988). One example of this can be seen today in the Singapore Hokkien word *suka* meaning ‘like’ or ‘enjoy’, that is a loanword from Bahasa Melayu.

The promotion of Mandarin was supported with language lessons that focused on the use of Standard Mandarin, while correcting deviations from ‘standard’ pronunciations (Dixon, 2005; Lock 1988). To achieve this, the use of

Hanyu Pinyin was introduced as a pronunciation guide, and Mandarin teachers were re-trained on new Standard Mandarin teaching materials (Lock 1988). Slogans for the campaigns over the years have also framed the use of Mandarin in a positive light, in opposition to the use of ‘dialects’ with slogans such as “Mandarin’s In. Dialect’s Out.”, and “If you’re a Chinese, Make a Statement – In Mandarin.”. In the early years of its inception, the Speak Mandarin campaign was intended to foster in-group cohesion between different Chinese ‘dialect’ groups, while also serving as a linguistic anchor for the transmission of ‘Chinese cultural values’ in the face of perceived negative “western” influences (Xie & Cavallaro, 2016). The success of the campaign in this regard is contested, as studies suggest that the decline of vernacular Chinese home languages instead resulted in a corresponding loss of dialect group-based Chinese practices and values which are typically transmitted between generations via oral tradition (Teo, 2005). Later iterations of the campaign in the 1990s then began to place emphasis on the importance of Mandarin proficiency for business and commerce dealings with China (Ng, 2014; Teo, 2025). The Speak Mandarin Campaign has continued to be held annually every September, and its overall effectiveness in promoting Mandarin as a lingua franca for the Chinese population of Singapore can be seen in an increase in the use of Mandarin, and a decrease in the use of Chinese vernacular languages over the years. While only 0.1% of the population identified as Mandarin speakers in 1957, 24.7% of Chinese Singaporeans reported using Mandarin as a home language in 1980 (Lock, 1988), and 64% of Chinese Singaporeans surveyed in 1996 reported using Mandarin in daily life, with only 34% reported using other vernacular Chinese languages (Xu et al., 1999 as cited in Dixon, 2005). Attitudes towards Mandarin have also shifted such that the majority of Singapore Chinese respondents in recent surveys now identify Mandarin as their ‘Mother Tongue’ (Ng, 2014; Starr & Hiramoto, 2019).

Thus, language policies have shaped the current pattern of ‘English-knowing’ bilingualism (Pakir, 1993; 2008) that is prevalent in the country. In the 2020 census, 48.3% of residents reported speaking English as one of their primary home languages (DOS, 2021). The use of other vernacular languages has been largely replaced with official ‘Mother Tongue’ languages. As discussed, Mandarin Chinese is now the most commonly spoken Chinese variety in Singapore, with the percentage of speakers of other Chinese vernacular varieties dropping from 76.2% in 1980 to 8.7% in 2020 (DOS, 2021). Some local language varieties are also in danger of extinction – such as in the case of Baba Malay whose ethnic speakers identify as Peranakan, but are designated as ethnically Chinese in National documents, and allocated Mandarin Chinese as their official ‘Mother Tongue’ language (Cavallaro & Ng, 2014). The language repertoires of individuals have also shifted away from a polyglossia of multiple languages, with the majority of individuals in the country currently reporting literacy in only two to three languages (DOS, 2021). In the 2020 census, 45.8% of residents reported being bilingual in English and Mandarin, 13.5% reported being bilingual in English and Malay, and 3.6% reported being bilingual in English and Tamil (DOS, 2021).

1.2.2 English-Mandarin bilingualism in Singapore

1.2.2.1 Overview of English-Mandarin bilingualism in Singapore

The current population of Singapore English-Mandarin bilinguals presents us with a largely homogenous group of bilinguals. As primary and secondary school education is compulsory in Singapore, the majority of English-Mandarin bilinguals in Singapore are exposed to at least 10 years of formal bilingual education (i.e., formal academic instruction in English with ‘Mother Tongue’ language classes) between the ages of 7 to 16 years. Despite this, there nevertheless remains a degree of linguistic heterogeneity within our target group of bilinguals. This can be particularly observed in differences in proficiency in Mandarin Chinese, as well as in familiarity with different varieties of Mandarin Chinese.

Current literature suggests that Mandarin proficiency amongst Singaporeans is strongly reliant on intergenerational transmission of the language, as Mandarin use in schools is largely confined to the context of a limited number of second language classes a week (Curdt-Christiansen & Sun, 2016; Xie & Cavallaro, 2016). According to the cognitive maturity model of language acquisition, individuals are more likely to attain higher, native-level, proficiency in languages that are acquired early in life during the sensitive period when exposure to different speech sounds shapes the categorical and structural manner in which languages are perceived and produced (Hammarberg, 2014). Current literature remains divided on how long this sensitive period lasts for, but research has suggested that the cut-off for high proficiency native-level language acquisition ranges from around 3 years of age to 7 years of age (Hammarberg, 2014). As a result of this, individuals from primarily Mandarin-speaking households who are exposed to Mandarin use from birth are likely to have higher proficiency and greater sensitivity to the structural regularities of Mandarin, as compared to individuals from primarily English-speaking households who may only be exposed to Mandarin later in childhood and only during formal language classes in school.

This difference in home language use has historically presented itself as a linguistic divide within the local ethnic Chinese community, and can be seen in the fairly equal enrolments of Singapore Chinese students in English-medium and Mandarin-medium schools up to 1959 (Lee & Phua, 2020), prior to increases in preferences for English-medium schools which led to the eventual mandatory use of English for instruction in all public schools (Dixon, 2005; Lee & Phua, 2020). Attendance at each type of school contributed to a linguistic divide between “English-educated” and “Chinese-educated” communities, with “English-educated” individuals primarily speaking English as their main home language, and “Chinese-educated” students primarily speaking Mandarin as their main home language (Chew, 2013). The intergenerational influence of this linguistic divide can still be seen in the 2020 census of Singapore which shows that 30% of ethnically Chinese residents report using Mandarin Chinese as their dominant home language, while 48% report using English as their dominant home language (DOS, 2021). Thus, it is likely that we will find a variety of levels of Mandarin proficiency amongst the local population of young adult English-Mandarin bilinguals.

1.2.2.2 Varieties of Mandarin Chinese in Singapore

Among the Mandarin Chinese spoken in Singapore, there are likely to be differences in the varieties of Mandarin Chinese that English-Mandarin bilinguals in Singapore are familiar with. The standard variety of Mandarin Chinese that is currently taught in formal education settings in Singapore is known as Putonghua, and is based on a variety of Mandarin that historically originates from the Northern and South-Western regions of China (Francis, 2016). Putonghua, or Standard Singapore Mandarin, shares phonological similarities to Beijinghua spoken in contemporary Beijing and was primarily introduced to Singapore as a novel second language for a large percentage of the population (as discussed in the previous section). As the majority of the ethnic Chinese population in Singapore are descended from seafaring migrants from the South-Eastern region of China, the Chinese languages that were originally brought to the country consisted of Southern Min Chinese languages which are mutually unintelligible with Mandarin (Li & Thompson, 1981; Starr & Wang, 2021). Thus, the introduction of Mandarin to serve as a common lingua franca within the Singapore Chinese community resulted in a colloquial variety of Mandarin that features heavy influence from these vernacular Southern Min Chinese languages. As a result of this, contemporary Singapore Mandarin often features individual variations in phonology that draw different degrees of influence from the ‘formal’ variety of Standard Singapore Mandarin, and the vernacular Southern Min influenced variety of colloquial Singapore Mandarin.

The influence of Southern Min languages on Singapore Mandarin can be seen in the phonology of Mandarin spoken in Singapore. For instance, documentations of Singapore Mandarin in the 1980s and in more recent years show that Singapore Mandarin speakers typically do not produce rhotacization in speech that is typical in standard Mainland Mandarin (also known as ‘er-hua’), due to the perception that it functions as a sociolinguistic marker of speech exclusive to Mainland Mandarin speakers (Starr, 2022). Similarly, it is also common for Singapore Mandarin speakers to produce Mandarin palatal frication /ɕ/ as the alveolar /s/ due to the lack of a palatal /ɕ/ in the phonology of Hokkien (Starr, 2022; Starr & Wang, 2021). However, one clear example of this Southern Min influence that has likely begun to change over the years can be seen in the presence of the Mandarin alveolar-retroflex phoneme contrast which we will focus on in this project. Documentations of Singapore Mandarin in the 1980’s suggest that the local variety lacks a clear alveolar-retroflex phoneme contrast due to the influence of languages such as Hokkien and Teochew which do not feature a retroflex place of articulation. For instance, observational studies conducted by Lock (1989) and Ng (1985) found that there was a large degree of merger of the alveolar (e.g., /s/, /ʈs/, and /ʈs^{h}/) and retroflex (e.g., /ʂ/, /ʈʂ/, and /ʈʂ^{h}/) phonemes in Singapore Mandarin, with the retroflex phoneme often being produced as alveolar. This sets the local Singapore variety of Mandarin apart from the standard variety of Beijing

Mandarin which has a clear alveolar-retroflex phoneme contrast (e.g., Chang, 2012). However, recent studies suggest that younger speakers of Singapore Mandarin may be starting to produce the alveolar-retroflex contrast more clearly in speech. For instance, Chen et al., (2016) found that primary school-aged children displayed more consistent use of the retroflex feature in their production of Mandarin speech as compared to older generations of Singapore Mandarin speakers. Similarly, using a word reading task, Starr (2022) found that only a minority of Singapore Mandarin speaking children produced retroflex sibilants as alveolar instead. This shift is likely related to the increase in the levels of exposure that young Singapore Mandarin speakers have to the standard variety of Mandarin in formal education (i.e., at school as well as in extracurricular language enrichment centres), and through mass media in standard Singapore and Mainland Mandarin (Starr & Kapoor, 2021). In their recent studies, Starr and colleagues have also suggested that the increase in the use of the Mandarin alveolar-retroflex contrast may be linked to the reduction in exposure to the vernacular Chinese home languages that were common only a few decades ago, coupled with the increasing influence of English as a home language which contains similar post-alveolar phoneme contrasts (e.g., /s/-/ʃ/, Starr, 2022; Starr & Wang, 2021). At the same time, young adult Singaporeans' perceptions of standard Mandarin have also begun to change from perceiving the alveolar-retroflex contrast as an undesirable "foreign" feature, to a marker of a more desirable, high prestige variety of Mandarin (Chong & Tan, 2013; Lock 1988).

Thus, we can see that the local population of young adult Singapore English-Mandarin bilinguals comprise a largely homogenous group of bilinguals in terms of the languages that they are exposed to in formal education and at home (i.e., English and Mandarin). At the same time, individuals within this group of bilinguals also have heterogenous levels of proficiency in their known languages, particularly the different levels of proficiency they have for the varieties of Mandarin that are spoken in Singapore as a result of varying levels of exposure to different varieties of Mandarin and other Chinese languages at home. This presents us with a suitable target population *within* which we can examine the language-specific account of bilingual learning advantages. If higher levels of Mandarin proficiency are linked to language-specific learning advantages, we expect that English-Mandarin bilingual adults with higher levels of Mandarin understanding ability should show stronger learning effects for stimuli that share structural features with the speech sounds in Mandarin. We therefore set out to test this hypothesis in the local population of English-Mandarin bilingual adults. We chose to use a distributional learning paradigm in this investigation as studies have shown that distributional learning can be utilised as a robust method of identifying learning advantages tied to language experiences in adults (Chládková & Šimáčková, 2021; Chládková et al. 2022). Moreover, distributional learning paradigms are easily adapted for participants across the lifespan (e.g., Infants: Maye et al., 2022; Children: Vandermosten et al., 2018), making such a paradigm well-suited for potential studies with younger participants.

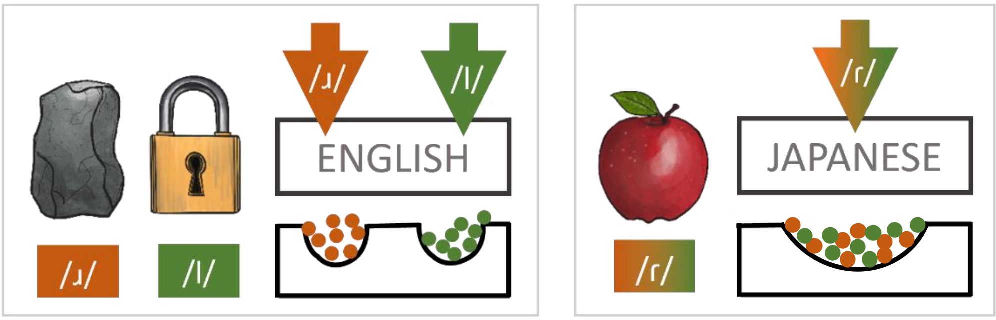
1.3. Distributional Learning

1.3.1 Distributional learning in early language acquisition

Distributional learning refers to the process by which meaningful patterns of statistical regularities are extracted from sensory input. This process occurs rapidly in young children, particularly in early infancy (Gómez, 2017), and plays a key role in the acquisition of structural features of languages of exposure early in life. Through distributional learning, infants acquire phoneme categories that are tuned to the unique featural characteristics of their native languages. Such learning follows exposure to high frequencies of meaningfully contrastive speech sounds early in life (Werker et al., 2012). According to Jusczyk (1993), listening experience plays a key role in modifying the acoustic-phonetic space of speech perception, thus changing how similarly or differently different speech sounds are perceived by individuals based on their unique language background. As a result of this, individuals who hear a high frequency of certain speech sounds distributed differently across acoustic space (e.g., a bimodal distribution) will typically develop a perceptual sensitivity tuned to identifying these sounds as different phonemes, while individuals who are exposed to a language environment in which these same speech sounds are not differentiated across acoustic space (e.g., a flat or unimodal distribution) will typically develop a perceptual sensitivity that is not as finely tuned to identifying differences between these sounds (see Figure 1.2 for illustration of this effect).

Figure 1.2

Visual example of modification of acoustic-phonetic space in speech perception following listening experience for English (left) as compared to Japanese (right).



Note. /ɹ/: voiced alveolar approximant; /l/: voiced alveolar lateral approximant; /ɾ/: voiced alveolar tap

Examples of the acquisition of native-language phoneme sensitivities following early-life language exposure can be seen in observational studies of infants. While young infants (under 8 months of age) generally begin life sensitive to a wide range of varied speech sound contrasts from different languages (Werker & Tees, 1984), the influence of exposure to native speech eventually leads to older infants developing phoneme category sensitivities that are more closely aligned with phonemic variations typical to their native language. An example of this can be found in a study by Werker and Tees (1984) which established that Hindi and English-hearing infants were equally sensitive at distinguishing the Hindi dental /d̪/ and retroflex /ɖ/ stop consonant phonemes at 6 months-of-age. However, this sensitivity diminished at 10-12 months-of-age for English-hearing infants, while Hindi-hearing infants retained sensitivity to this phoneme contrast. As /d̪/-/ɖ/ are contrastive in Hindi speech but not English, Werker and Tees (1984) proposed that Hindi-hearing infants would have been exposed to a higher frequency of speech tokens that fall distinctly into the two separate categories, leading to maintenance of the perceptual sensitivity and consolidation of acoustic space into two separate phonemes. Conversely, the English-hearing infants would not have been exposed to this same distinct distribution of meaningfully contrastive speech sounds, thus leading to an attenuation of the perceptual boundary between the two speech sounds, resulting in a single phonemic representation /d/ for this acoustic space.

Similarly, Tsushima et al. (1994) tested Japanese-hearing infants on their ability to discriminate /ɹ/ and /l/ – contrastive phonemes in English but not Japanese. They found that while infants between the ages of 6-8 months could reliably discriminate between the two phonemes, infants who had been exposed to Japanese for longer (10-12 months) could no longer do so. At the same time, Narayan et al. (2010) found that while *neither* English- nor Tagalog-hearing 6-to-8-month-old infants were capable of discriminating between the perceptually complex /n/-/ŋ/ contrast found in Tagalog, 10-12 month-old Tagalog hearing infants *were* able to discriminate this contrast while the 10-12-month-old English-hearing infants were not, thus showing that listening experience in Tagalog led to development of two phoneme categories for the Tagalog-hearing infants but not for the English-hearing infants at this age.

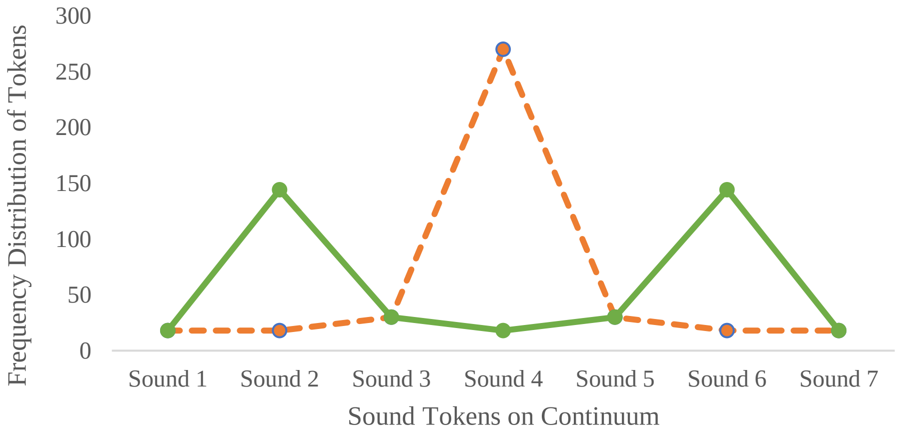
Listening experience has also been demonstrated to contribute to the rate at which infants *lose* contrast sensitivity to non-native phoneme contrasts. For instance, Anderson et al. (2003) tested English-hearing 6.5-month and 8.5-month-olds on non-native coronal stop contrasts, and non-native dorsal stop contrasts. While both the 6.5-month and 8.5-month-olds in this study were able to discriminate between the non-native dorsal stop contrast, only the 6.5-month-olds were able to discriminate between the non-native coronal stop contrast. This was suggested to have been the result of high frequencies of coronal stop contrasts in English leading to an earlier loss of non-native coronal stop contrast as compared to the dorsal stop contrasts which are less frequently encountered in English. These studies show that language listening experience early in life plays a crucial role in the development and consolidation of native language phoneme categories through distributional learning. At the same time, existing research has established that distributional learning is not limited to naturalistic settings early in life, and can even be elicited in a controlled experimental context, thus making distributional training a suitable paradigm for our investigation into the language-specific bilingual learning advantage.

1.3.2 Distributional learning in experimental contexts

Similar to distributional learning in naturalistic settings, experimental studies have shown that individuals are similarly able to extract statistical information from different frequency distributions of speech sounds following listening experience. The nature of speech sounds is typically continuous rather than strictly categorical, and the range of acoustic outcomes between different speech sounds form a speech sound continuum. One frequently used method of studying distributional learning in a lab-based setting is to examine if a change in sensitivity to speech sound boundaries can be obtained in participants following training on speech sounds presented with a bimodal or unimodal distribution across acoustic space. In a bimodally distributed training set, a larger number of sound tokens are drawn from tokens closer to the two endpoints of a speech continuum. In a unimodal distribution, the majority of sounds in a training set are obtained from the midpoint of the speech continuum (see Figure 1.3 for example of bimodal and unimodal distribution of sound tokens).

Figure 1.3

Example of unimodal (orange, dashed line) and bimodal (green, solid line) distributions of speech tokens presented in an experimental paradigm.



Studies have demonstrated that bimodal training typically results in enhanced sensitivity to contrastive phoneme categories, while unimodal training typically results in a perceptual collapse across these categories. This mirrors the pattern of native language phoneme acquisition typically observed in infants early in life following extensive exposure to native language speech sounds. By comparing the perceptual sensitivity of participants trained on bimodally distributed stimuli to their pre-training performance, or the performance of participants trained on unimodally distributed or unrelated control stimuli, studies can assess how distributional training recalibrates participants' phoneme contrast sensitivities. This distributional training paradigm has been used with infants (e.g., Maye et al., 2022; Wanrooij et al., 2014), children (Vandermosten et al., 2018), and adults (e.g., Chládková & Šimáčková, 2021) with phoneme contrasts based on voice onset time (Maye et al., 2002; Yoshida et al., 2010), formants (Wanrooij et al., 2014) and tonality (Liu & Kager, 2017). Crucially, distributional learning studies can be used to elicit learning effects for perceptual sensitivity to fine grained speech stimuli that are perceptually difficult to tell apart – thus making it an ideal paradigm for assessing genuine learning effects.

1.3.2.1 Distributional learning in infants

Studies have shown that infants have a particularly robust ability to learn from distributional training. In fact, evidence for learning effects following distributional training can be seen in infants as young as 2 months-of-age following as little as two minutes of training (e.g., Saffran et al., 1996). Evidence of such learning occurring can be found in both neuroimaging as well as behavioural studies. For instance, Wanrooij et al. (2014) conducted an ERP study in which they trained 2 to 3-month-old Dutch infants on either a bimodal or unimodal distribution of sounds on a novel /ɛ/- /æ/ vowel contrast. Using an oddball paradigm to determine if the infant participants could distinguish between the vowel sounds, the authors found that bimodally trained infants had significantly larger mismatch response amplitudes to oddball /æ/ stimuli than unimodally trained infants, indicating a higher sensitivity to the two categories of vowels as compared to unimodally trained infants. This showed that the type of training the infants underwent had an effect on how they perceived the sounds they were trained on, with the bimodally trained infants acquiring a perceptual sensitivity in which they more clearly perceived the /ɛ/ and /æ/ test tokens as belonging to two phonemic categories. On the other hand, the listening experience of the unimodally trained infants led them to acquire a less pronounced perceptual sensitivity for the /ɛ/ and /æ/ test tokens as compared to the bimodally trained infants.

Behavioural studies with infants can be used to obtain similar distributional learning effects. For instance, Maye et al. (2002) trained 6- and 8-month-old infants on either a unimodal or bimodal frequency of tokens from an 8-step /da/-/ta/ continuum in which the sound tokens differed incrementally in small acoustic steps from each other in VOT. Using a novelty preference looking paradigm, they then tested both the unimodally and bimodally trained infants on their perception of /da/ and /ta/ sound tokens. As expected, the results of the study showed that bimodally trained infants showed evidence of having acquired a two-category perceptual sensitivity for /da/ and /ta/, as though they belonged to two separate phonemic categories. On the other hand, the unimodally trained infants in the study did not show evidence of acquiring this two-category perceptual sensitivity, responding to the test /da/ and /ta/ sound tokens on the continuum as though they were perceptually similar. Similarly, Yoshida et al. (2010) conducted a distributional learning paradigm on 10- to 11-month-old infants with stimuli similar to that of Maye et al. (2002), and found that bimodally-trained older infants likewise showed a stronger degree of perceptual sensitivity to the /da/ and /ta/ test tokens as belonging to two phonemic categories as compared to the unimodally trained infants. However, while the results of Yoshida et al. (2010) mirrored that of Maye et al. (2002), a longer training period was needed for these older infants to demonstrate any learning effects, likely due to the fact that these older infants had already begun consolidating phonemic categories specific to those of their native languages.

1.3.2.2 Distributional learning later in life

Research has shown that the ability to learn novel contrasts from distributional cues declines with age as infants consolidate phoneme sensitivities tuned to their native language, with older infants showing less sensitivity to distributional cues. This becomes most apparent at around 12 months of age, with individuals continuing to lose sensitivity for contrastive perception of non-native speech sound contrasts with as they age (Werker & Tees, 1984; Reh et al., 2021). However, distributional learning effects can still be elicited in children following an increased amount of training time. Using a longer training period of 10 minutes, Vandermosten et al. (2018) trained Grade

3 Dutch children on either a unimodal or a bimodal distribution of sound tokens drawn from a novel 8-step Hindi dental-retroflex contrast continuum in which sounds ranged in incremental degrees of “dental-ness” or “retroflex-ness” from a dental /d̪a/ to a retroflex /ɖa/. They tested the participants on their sensitivity to the categorical identify of each sound using a 2-Alternative Forced Choice (2AFC) paradigm both before and after the training period, and compared their results at each stage. The results of this study revealed that neurotypical children who had been trained on a bimodal distribution showed significantly larger increases in phoneme contrast sensitivity at the post-training test as compared to the unimodally trained participants. Therefore, despite the fact that individuals show a decrease in the ability to access statistical information from distributional frequencies of unfamiliar speech sounds with age, this study shows that the effects of bimodal distributional training can still be observed in children after a relatively short period of training.

However, distributional training studies on adults tend to have results that are not as clear as in studies of infants and children. A distributional training ERP MMN study by Wanrooij et al. (2014) compared the results of Dutch 2- to 3-month-old infants and adults, and found that bimodal training on English /æ/ and /e/ vowels only resulted in learning effects for infants, but not for adults. Terry et al. (2015) found that their Australian English-speaking adult participants were so poor at extracting information from distributional training stimuli that they were unable to demonstrate any significant learning effects following bimodal training on a novel Dutch /ɑ/-/a:/ vowel contrast, even when the stimuli were altered to enhance differences in the two vowel sounds. In fact, an ERP MMN study by Liu et al. (2022) found that Australian English-speaking participants also showed *poorer* discrimination of a Mandarin tonal contrast following bimodal training, while unimodally trained participants showed no difference in their pre- and post-training neural responses. At the same time, studies such as that of Hayes-Harb (2007) and Barrios et al. (2022) have shown that while adult participants may still be able to extract an amount of phoneme category information from bimodal distributional learning stimuli, they may rely more heavily on different learning strategies, such as identifying lexical cues to consolidate information on novel contrasts. Thus, adults appear to be weaker than infants and children at accessing the mechanism needed to extract information from distributional cues in novel speech sounds.

On the other hand, some studies have nevertheless found that adults can successfully show learning effects following distributional training on a bimodal distribution of speech sounds. This can be seen in a series of three studies by Escudero and colleagues who trained Dutch-learning, Spanish-speaking participants on the Dutch /ɑ/-/aː/ vowel contrast. Escudero et al. (2011) trained their participants on a naturalistic or enhanced (exaggeration in F1 and F2) bimodal distribution of sounds on the Dutch /ɑ/-/aː/ vowel contrast, and found that both the natural and enhanced condition resulted in significantly improved contrast sensitivity at post- compared to pre-training, but participants in the enhanced condition showed greater improvement. Impressively, Escudero and Williams (2014) found that similar learning effects could be found in bimodally-trained adult language learners after as little as two minutes of training time, and that such learning effects were long-lasting and could still be observed 12 months after training. Wanrooij et al. (2013) further outlined the nature of this learning effect by demonstrating that bimodal training on this set of stimuli resulted in Spanish-speaking participants not only improving in their overall perception of the /ɑ/-/aː/ vowel contrast, but also in their ability to access more acoustic cues (e.g., differences in duration, F0, F1, F2, F3) embedded in the sound tokens for phoneme discrimination.

Similarly, studies by Chládková and colleagues have shown that distributional learning effects can be observed in some adults following bimodal training. Chládková and Šimáčková (2021) trained Czech- and Greek-speaking adults on a bimodal or unimodal distribution of durationally-cued vowel sounds on a novel /ɑ/-/aː/ continuum. They found that the Czech-speaking adults who had experience with durational cues in phoneme contrasts in Czech showed significant learning effects for this novel contrast following bimodal training, while Greek-speakers who did not have any experience with durational cues in phoneme contrasts did not. In their ERP MMN study on Spanish speakers, Chládková et al. (2022) found that bimodal distributional training could lead to a shift in the location of the category boundary of their participants' perception of the Spanish /i/-/e/ vowel contrast. In particular, they found that participants showed significantly larger learning effects when they were trained on /i/-/e/ stimuli that utlitised F1 variation as an acoustic cue, rather than /i/-/e/ stimuli that utilised durational variation as an acoustic cue. As

F1 variation is an acoustic cue for phoneme contrasts in Spanish while duration variation is not, this study appears to show that adults are particularly able to learn from distributional training if the training stimuli shares featural characteristics with their existing native language phoneme repertoire, through transfer of knowledge from their native language.

Crucially, one key difference between adult studies that have and have not found distributional learning effects appears to be the amount of overlap present between structural regularities in the to-be-learnt contrast, and the languages in the participants' repertoires. For instance, the study by Chládková and Šimáčková (2021) trained participants on a synthesised vowel contrast based on Czech-like durational cues, and found that only the Czech-speaking participants showed learning effects for this contrast following bimodal training. Similarly, the study by Chládková et al. (2022) showed that their Spanish-speaking adult participants were best able to learn a vowel contrast when it contained structural F1 regularities similar to the regularities found in their native language. In the case of Escudero and colleagues, the participants had already begun learning Dutch or were living in the Netherlands, and would therefore have had relevant prior exposure to the Dutch /ɑ/-/aː/ vowel contrast that they were trained on in the studies (Escudero et al., 2011; Escudero and Williams, 2014; Wanrooij et al., 2013). On the other hand, adults who were trained on completely novel speech sounds which did not have any structural overlaps with the languages in their repertoires appear to have been unable to utilise distributional cues for learning, instead focusing on more salient cues at hand such as acoustic cues (e.g., Liu et al. 2022), and lexical cues (Hayes-Harb, 2007; Barrios et al., 2022).

Therefore, we can see that distributional learning can occur in adults if there is sufficient overlap between the structural regularities of their known languages, and the to-be-learnt language stimuli. This supports the idea of a language-specific transfer advantage as discussed in Section 1.1.3, and shows that a distributional learning paradigm will be well-suited for our investigation of language-specific transfer effects within our target population of Singapore English-Mandarin bilinguals.

Chapter 2: Investigating a language-specific learning advantage with a Hindi dental-retroflex stop contrast (distributional learning)

2.1 Introduction

To begin this project, we planned an investigation into the validity of a language-specific transfer advantage in bilinguals. To do this, we set out to determine if we could find evidence for such learning advantages in young adult Singapore English-Mandarin bilinguals with a distributional learning task featuring a novel Hindi dental-retroflex contrast. For the purposes of this thesis, the term 'distributional learning effects' will be used to refer to changes in perception of a phoneme contrast following training on a bimodally distributed frequency of sounds drawn from a continuum based on the target contrast. Studies assessing learning effects have generally utilised a separate control group to account for extraneous factors that might induce change. However, we were primarily interested in the nature of how language backgrounds interact with pre- and post-training changes in perception, rather than the precise underpinnings of the change in question. As such, we focused our analysis on the pre- and post-training changes in perceptual sensitivity within a single group of participants. This is in line with how distributional learning has been described and examined in existing distributional learning studies using bimodally distributed frequencies of speech sound stimuli for distributional training (e.g., Chládková et al., 2022). In this chapter, the term 'distributional learning effects' will specifically refer to increases in perceptual sensitivity for ambiguity embedded in a synthesised Hindi dental-retroflex contrast continuum following bimodal distributional training on sounds drawn from the continuum.

2.1.1 Retroflex overlap in Hindi and Mandarin

The retroflex place of articulation is ostensibly present in both Hindi and standard Mandarin (stop consonants in Hindi, sibilants in Mandarin). As such, we reasoned that the presence of alveolar-retroflex contrasts in Mandarin and dental-retroflex contrasts in Hindi could serve as a source of linguistic overlap between the two languages, from which language-specific transfer effects could occur. For instance, English-Mandarin bilinguals in Singapore would have experience with the alveolar-retroflex sibilant contrast found in Mandarin (e.g., /s/-/ş/), but not with a Hindi dental-retroflex stop contrast of /d̪/-/ɖ/, making a Hindi dental-retroflex contrast a good training target to focus on.

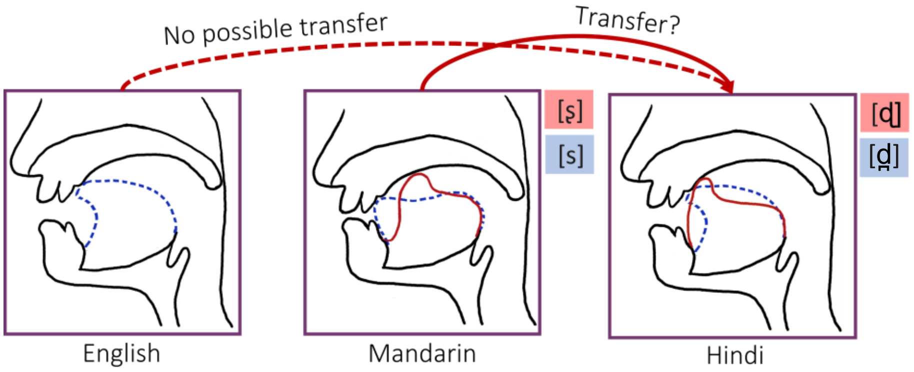
Additionally, English-Mandarin bilinguals in Singapore vary in their familiarity with standard Mandarin, and this may be reflected in their exposure to the Mandarin retroflex place of articulation. Unlike the standard Beijing variety of Mandarin which has a clear alveolar-retroflex contrast, varieties of Mandarin (e.g., Singapore and Taiwan Mandarin) which are more heavily influenced by Southern Min Chinese language varieties (e.g., Hokkien), tend to show degrees of phoneme contrast merger, or deretroflexion. In deretroflexion, retroflex phonemes may be produced with an alveolar place of articulation, or may be much less retroflexed in nature as compared to the standard variety of Mandarin (e.g., Chang & Shih, 2015).

Early studies of Singapore Mandarin in the 1980's show that Singapore Mandarin speakers tend to have different levels of deretroflexion in speech (Lock, 1989; Ng, 1985). At the same time, recent studies indicate that an increase in the perceived prestige of the standard variety of Mandarin appears to be resulting in an increase in the frequency of clear retroflexion in the speech of speakers of contemporary Singapore Mandarin (Chen et al., 2016; Chong & Tan, 2013). Familiarity with the feature of Mandarin retroflexion is likely to be associated with higher levels of Mandarin understanding abilities amongst contemporary Singapore Mandarin speakers as retroflexion is common in the standard variety of Mandarin used in the formal education system, as well as in commonly consumed forms of Chinese media where standard Mandarin articulation is favoured such as the news on television as well as the radio, and mass media from Mainland China (Chong & Tan, 2013; Starr, 2022; Starr & Kapoor, 2021). Mandarin proficiency in Singapore is strongly linked to intergenerational transmission of the language (Curdt-Christiansen & Sun, 2016; Xie & Cavallaro, 2016) such that individuals from primarily Mandarin-speaking households are more likely to have higher proficiencies in Mandarin as compared to individuals from primarily English-speaking households. As such, we reasoned that individuals with higher Mandarin understanding abilities would be more likely to be exposed to Mandarin retroflexion from these various sources of input across the lifetime, and from an earlier age as compared to individuals with lower Mandarin understanding abilities who would likely only have been exposed to Mandarin retroflexion for a more limited period of time during formal language classes introduced later in childhood (i.e., for a few hours a week between the ages of 7 to 16). We can utilise this source of intra-group linguistic variance to examine if differences in Mandarin understanding abilities may result in differences in distributional learning effects for a novel Hindi dental-retroflex contrast.

For this first study, we decided to conduct a replication of a distributional learning study with a Hindi dental-retroflex /d̪a/-/ɖa/ phoneme contrast continuum originally conducted by Vandermosten et al. (2018). We believed that this particular set of training stimuli would be learnable by our young adult Singapore English-Mandarin bilinguals, as this same set of stimuli have been used in learning studies on children and adult participants - albeit with a different learning paradigm with adult participants (e.g., Golestani et al., 2002; Golestani & Zatorre, 2009; Vandermosten et al., 2018). At this point of our study, we believed that the retroflex place of articulation in Hindi and Mandarin would have a sufficient amount of overlap to pursue our investigation into a language-specific transfer effect. However, further investigation into the nature of the Mandarin retroflex (particularly the Singapore Mandarin retroflex) following this study would suggest otherwise, as retroflexion in Singapore Mandarin is not a “true” retroflex and appears to approximate a more laminal post-alveolar place of articulation, as documented in Mainland Mandarin by Ladefoged and Wu (1984). Based on our initial understanding of the nature of Mandarin retroflexion, as well as the language-specific hypothesis of learning advantages in bilinguals, we reasoned that: 1) our participants would be able to show distributional learning effects in the form of increased perceptual sensitivity for phoneme contrast ambiguity embedded in a Hindi dental-retroflex contrast based on the featural overlap between Hindi and Mandarin retroflexion (depicted in Figure 2.1), and 2) we would find a positive correlation between distributional learning effects and Mandarin understanding proficiencies, as greater familiarity with the alveolar-retroflex contrast in Mandarin would result in larger transfer effects to a novel dental-retroflex contrast in Hindi. Despite our initial misconceptions about the lingual nature of Singapore Mandarin retroflex frication, we have nevertheless chosen to include this study in this thesis as it provides us with further insight into distributional learning paradigms tested with an adult population.

Figure 2.1

Proposed cross-language transfer effect based on shared feature of retroflexion between standard Mandarin and Hindi in Singapore English-Mandarin bilinguals.



2.1.2 Methodological adaptation of study

This study was a close methodological adaptation of the Hindi dental-retroflex distributional learning study conducted by Vandermosten et al. (2018) in which participants were trained on a 7-step Hindi dental-retroflex /d̪a/-/ɖa/ continuum with a distributional learning training phase, and distributional learning effects were assessed by comparing participants’ categorical perceptual sensitivity for the sound tokens at pre-training and post-training. We made four main changes to the methodology of the study to make it more suitable for examining the language-specific hypothesis of bilingual learning advantages. First, our study was conducted with young adult participants instead of children (as in the original study by Vandermosten et al., 2018). Second, we increased the number of sound tokens played during the bimodal distributional learning training phase to a total of 4020 sound tokens, while still maintaining the bimodal distributional frequency of sound tokens originally used by Vandermosten et al. (2018). This led to a training phase duration of around 30 minutes, a significant increase from the 8-to-10-minute training duration of Vandermosten et al. (2018). This modification was made as existing studies have shown that the amount of training typically needed for successful distributional learning increases with age. Third, rather than including a separate control group trained on a unimodal frequency distribution of sound tokens, we chose to assess distributional learning effects by comparing our participants’ pre-training responses to their post-training responses following training on a bimodal distributional frequency of sound tokens. Using a within-group comparison of pre vs. post-training responses following bimodal distributional training is an established method of identifying distributional learning effects in the form of changes in perception of phoneme contrast continua, such as in the case of Chládková et al., (2022) who examined distributional learning effects in adult participants following bimodal distributional training on different sets of phoneme contrast stimuli.

Finally, we altered the nature of the training phase from an attended auditory task to an unattended auditory task. In Vandermosten et al.’s (2018) study, the training phase included catch-trials in which participants responded to oddball sound tokens that were not drawn from the 7-step Hindi dental-retroflex /d̪a/-/ɖa/ continuum. In our study, participants were informed that they did not have to attend to, or respond to any of the sounds they heard. This choice was made as initial feedback from pilot study testers revealed that a high amount of working memory load was inadvertently recruited in adults with a similar catch-trial task, and testers reported difficulty in effortfully tracking different sound tokens across our much longer training phase. The ‘bottleneck’ theory of perceptual processing suggests that cognitive resources shared between different cognitive functions are limited, and are thus primarily focused on the most effortful task at hand (e.g., Broadbent, 1958) – in this case, the catch-trial task. We therefore reasoned that an unattended training task would free up valuable cognitive resources for the purposes of distributional learning rather than for completing the catch-trial task.

To promote open science practices, we have made all relevant data, task files, stimuli, and analysis files for this study available on the Open Science Framework (https://osf.io/agrpj/).

2.2 Pilot Study

Before conducting a full-scale study, we first needed to determine if bimodal distributional training on this novel Hindi dental-retroflex contrast continuum would elicit any distributional learning effects in the form of changes in perceptual sensitivity for ambiguity embedded in this Hindi dental-retroflex contrast continuum in young adult Singapore English-Mandarin participants. Thus, we conducted a small-scale pilot study to examine distributional learning effects.

2.2.1 Methods

2.2.1.1 Participants

16 participants (10 female, 6 male) were recruited from the student population of Nanyang Technological University in exchange for course credits. One male participant was excluded from the study due to colour blindness, as a key feature of our study relies on colour to differentiate visual stimuli. The remaining 15 participants were aged between 20-29 years of age (*Median* = 22). All participants self-identified as English-Mandarin bilinguals who had lived in Singapore for the majority of their lives. Informed consent was collected from all participants prior to commencement of the study, and

this study was approved by the IRB of Nanyang Technological University institution (IRB-2019-01-034).

2.2.1.2 Equipment

The distributional learning task was conducted on a 10-inch LCD screen on a Microsoft Surface Go computer. All stimuli were presented in OpenSesame (Mathôt et al., 2012). Audio stimuli were presented to the participants with a pair of Audio-Technica Over-ear Monitoring Headphones ATH-M40x.

2.2.1.3 Stimuli

Visual stimuli consisted of two cartoon aliens (one purple and one orange), and an illustration of a "transmission device" used to "send messages to aliens" (see Figure 2.3 for example). During the training phase, participants were also presented with muted cartoons (selected clips from Series 2 of Mr Bean: The Animated Series; Atkinson et al., 2003).

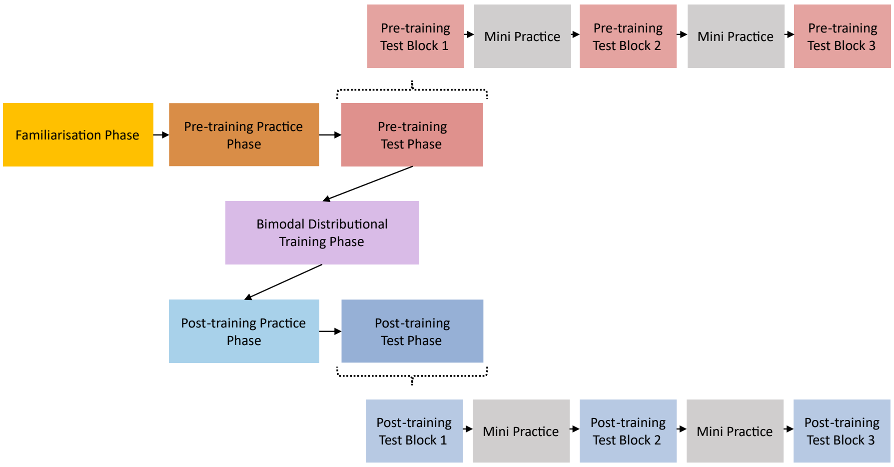
Audio stimuli were identical to that of Vandermosten et al. (2018) and consisted of a synthesised Hindi dental-retroflex /d̪a/-/ɖa/ stimulus continuum of seven sound tokens that were each 220ms in duration. Sound 1 in the continuum represented a dental /d̪a/ and Sound 7 represented a retroflex /ɖa/. Sounds 2 to 6 represented intermediate audio tokens with varying levels of dental-retroflex contrast ambiguity. The end points (Sounds 1 and 7) of the continuum were synthesised to be very close to the category boundary of native speakers of Hindi, and the initial consonant of each of the sound tokens on the continuum were synthesised to be evenly spaced from each other in terms of acoustic distance in the transition of formant 3 (in 111Hz steps) and the centre frequency of the burst (in 217Hz steps) (Vandermosten et al., 2018).

2.2.2 Procedure

The procedural flow of the study is shown in Figure 2.2.

Figure 2.2

Procedural flow of Hindi dental-retroflex distributional learning study.



2.2.2.1 Familiarisation phase

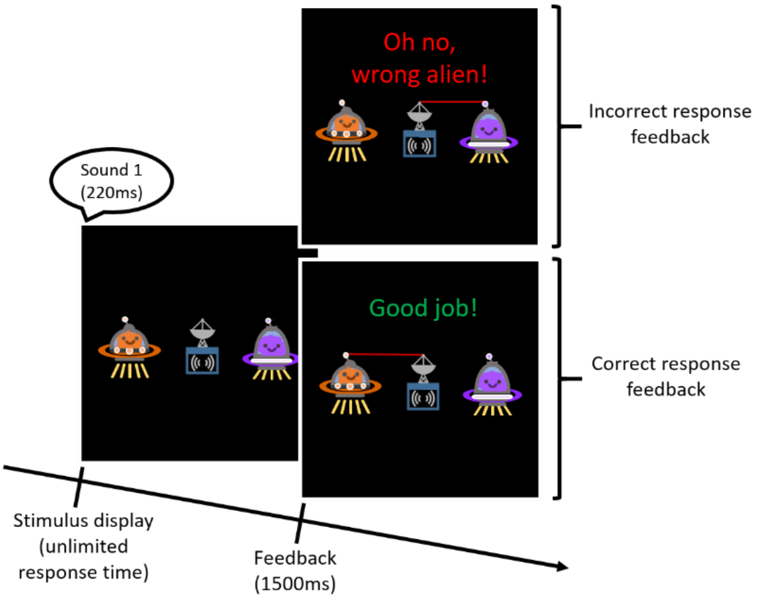
The familiarisation phase was presented at the start of the study procedure to introduce the participants to the two novel Hindi dental and retroflex speech sound categories, and to allow them to familiarise themselves with the sound-to-alien pairing used in the study. On each familiarisation trial, participants an image of one of the aliens was displayed at a random location on the screen while its corresponding sound token was played. The orange alien was always paired with Sound 1 (the dental /d̪/ end of the continuum), and the purple alien was always paired with Sound 7 (the retroflex /ɖ/ end of the continuum). Participants were presented with on familiarisation block in which each alien was presented a total of 10 times in alternating order (20 trials in total). Participants were informed that no responses were required during the familiarisation block.

2.2.2.2 Pre- and post-training practice phase

To ensure that participants were aware of the sound-to-alien pairing used in the study, they completed a practice block in which they were told that their job was to “send messages to aliens” during the sound-to-alien pairing in the familiarisation phase. On each practice trial, participants were presented with an orange alien and a purple alien on either side of the screen, and a “transmission device” in the middle of the screen. Either Sound 1 or Sound 7 was played, and participants were instructed to make a keypress under the alien the sound matched with in order to a send it a message. Following a keypress response, an animated beam would move from the “transmission device” to the selected alien, while a “transmission sound” played (created by Freesound.org contributor Jagadamba, downloaded from https://freesound.org/people/Jagadamba/sounds/253908/). For each practice trial response, participants were provided with written onscreen feedback. There was a total of eight trials in each block, with Sounds 1 and 7 played four times each in random order over the course of each block. Figure 2.3 shows an example schematic of the stimuli presented during a practice trial.

Figure 2.3

Example of stimuli presented during a practice trial.



Note. Not drawn to scale.

2.2.2.3 Pre- and post-training test phase

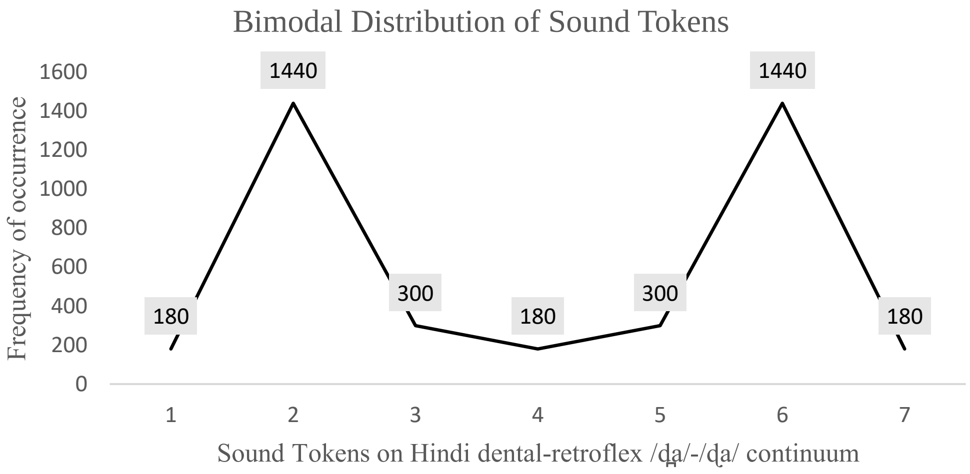
To establish a baseline for each participant, all participants completed three pre-training test blocks prior to the bimodal distributional training phase. Following the bimodal distributional training phase, participants then completed three post-training test blocks. The test phases were similar to the practice phases, with the exception that participants heard all seven of the sound tokens on the Hindi dental-retroflex /d̪a/-/ɖa/ continuum, and were not given any feedback following each response made. Each test block had a total of 14 trials. On each trial, the sound token played was randomly selected from the full 7-step Hindi dental-retroflex stimulus continuum. Each of the 7 sounds was played twice in random order over the course of each test block. To ensure that participants had sufficient time to make a response for each sound token heard, there was no response time limit on each test trial. To maximise participants' accuracy in remembering the correct sound-to-alien pairing across each test phase, participants also completed one mini practice block between each test block. The mini practice blocks were similar to the main practice block with the exception that participants only heard Sounds 1 and 7 (the continuum endpoint tokens) two times each.

2.2.2.4 Bimodal distributional training phase

For the bimodal distributional training phase, participants were instructed to watch a muted cartoon while audio stimuli played. Participants were informed that no response would be needed from them during the training phase. During the unattended training paradigm, a total of 4,020 sound tokens were presented to participants in a random order, with an 80ms ISI between each sound. To achieve a bimodal frequency stimulus distribution, each of the 7 sounds from the stimulus continuum was played according to the frequency distribution shown in Figure 2.4. The training phase lasted around 30 minutes in total.

Figure 2.4

Frequency of sound tokens in bimodal distribution used for training.



2.2.2.5 Language ID

To obtain information on our participants' language repertoires, participants were asked to rate their ability in understanding the four official languages of Singapore (English, Mandarin, Malay, and Tamil), along with any other languages and dialects they understood on a Likert scale of 0, "I do not understand this language", to 7, "I have native-level understanding of this language". The language ID was administered to each participant following completion of the Hindi dental-retroflex distributional learning task.

2.2.3 Analysis plan

All analyses were conducted with R (R Core Team, 2020). In order to analyse our participants' responses in the pilot study, we obtained the number of purple alien (Sound 7 retroflex /ɖa/) responses each participant made for each of the seven tokens on the stimulus continuum at the pre-training and post-training test phases. We then computed the proportion of Sound 7 /ɖa/ responses out of all responses that were made for each of the seven sound tokens at each test phase separately. Pre-training and post-training psychometric slopes were fitted to this data for each participant separately, and decision gradient slope values of each psychometric curve were obtained using the *quickpsy* function in R (Linares & López-Moliner, 2016; R Core Team, 2020).

The decision gradient slopes obtained informed us about the participants' perceptual sensitivity for the Hindi dental-retroflex contrast at pre-training and post-training test phase. Steeper decision gradient slopes (and higher slope values) represent a more systematic pattern of responses, indicating greater sensitivity to the differences in dental-retroflex ambiguity between each sound token on the dental-retroflex continuum. On the other hand, flat slopes represent chance responding, and less perceptual sensitivity to dental-retroflex ambiguity between each sound token.

2.2.3.1 Pre-training control analysis

To ensure that exposure to the sound tokens presented prior to the bimodal distributional training phase had not led to any distributional learning effects, we first conducted a control analysis on the slope values of the first and third pre-training blocks to determine if significant changes in slope value had occurred prior to the training phase. Individual slope values were obtained separately for each participant for the first and third pre-training test blocks for comparison. As the sample size was small, a non-parametric Wilcoxon signed rank test was conducted to compare changes in slope values between the first and third pre-training blocks using the *wilcox.test* function from the *stats* package (R Core Team, 2020).

2.2.3.2 Distributional learning analysis

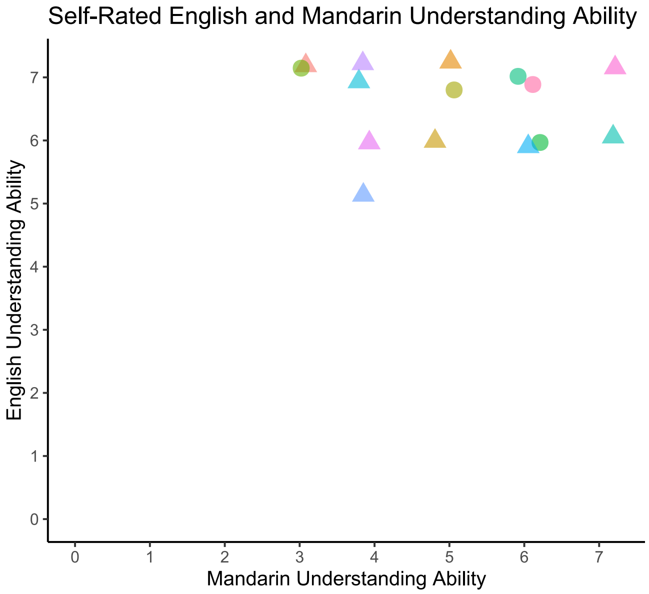
We then conducted our main analysis and assessed for distributional learning effects by comparing our participants' pre-training baseline slope values to their post-training slope values. As the sample size was small, a non-parametric Wilcoxon signed-rank test was carried out to test for difference in the participants' decision gradient slope values at pre- and post-training test using the *wilcox.test* function from the *stats* package (R Core Team, 2020).

2.2.4 Results

2.2.4.1 Language ID

Figure 2.5

Pilot participants' self-reported English and Mandarin understanding ability.



Note. *N* = 15. Circles: female, triangles: male. Responses jittered by .25 for visualisation.

Figure 2.5 shows the distribution of our participants' self-rated English and Mandarin understanding abilities. Overall, our participants' self-reported English understanding scores were high (Range: 5-7, *M* = 6.5, *SD* = 0.6), this indicating that they would not have had any issues in understanding the task instructions of the study. On the other hand, their Mandarin understanding scores showed more variability (Range: 3-7, *M* = 5.0, *SD* = 1.3). No participants reported being able to understand Hindi.

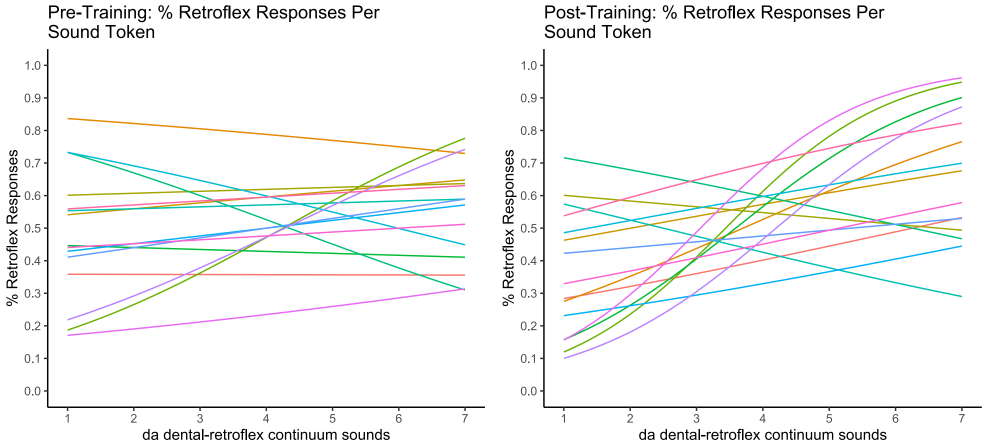
2.2.4.2 Pre-training control analysis

The control analysis revealed that there was no significant difference in decision gradient slope value between the first and third pre-training blocks, (*Z* = -.03, *p* = .98, *r* = -.008; pre-1: slope *M* = 0.99, *SD* = 3.80; pre-3 slope: *M* = 0.60, *SD* = 2.13). This indicates that no learning had occurred prior to bimodal distributional training.

2.2.4.3 Distributional learning analysis

Figure 2.6

Decision gradient slopes for individual participants at pre- and post-training test phase.



Note. *N* =15. Pre-training test phase (left). Post-training test phase (right)

Fitted psychometric slopes for the pre-training and post-training test phases are presented in Figure 2.6. From visual inspection, it is possible to observe that several participants had steeper slopes at the post-training test phase as compared to the pre-training test phase. The corresponding Wilcoxon signed-rank test revealed a significant difference in the pre-training and post-training slope values, ($Z = -2.61$, $p = 0.004$, $r = -.67$; pre- training slope: $M = 0.5$, $SD = 0.19$; post- training slope: $M = 0.27$, $SD = 0.33$). This indicates that learning had occurred during the bimodal distributional training phase.

2.2.5 Discussion

The results of the pilot study showed that a significant overall distributional learning effect could be obtained within a small sample of

Singapore English-Mandarin bilingual adults with our adaptation of Vandermosten et al.’s (2018) study. This suggests that the paradigm and training stimuli would be suitable for assessing distributional learning effects in a larger, more representative sample size of Singapore English-Mandarin bilingual adults. Moreover, we found a range of distributional learning effects, suggesting that this study paradigm would be suitable for quantifying individual differences in distributional learning effects. We also found a range of Mandarin understanding ability scores, suggesting that we could utilise this intragroup difference to examine transfer effects in a larger, more representative sample size. We thus preregistered a full-scale study based on the findings of this pilot study (Goh et al., 2020; https://osf.io/s6vdn).

2.3 Main Study

Following the pilot study, we carried out a full-scale study to investigate the hypothesis of language-specific transfer effects. Familiarity with the feature of Mandarin retroflexion is likely to be associated with higher levels of Mandarin understanding, as retroflexion is common in the standard variety of Mandarin typically used in formal education in Singapore (Chong & Tan, 2013). Therefore, we reasoned that if learning advantages in bilinguals can indeed arise from language-specific transfer effects, larger distributional learning effects should be found in individuals with higher levels of Mandarin understanding abilities.

To ensure that this main study would be well-powered, we conducted an a priori power analysis in G*Power 3.1.9.7 to determine an appropriate minimum sample size. The analysis revealed that a minimum sample size of $N = 30$ would be needed to observe an effect size equal to the size observed in the pilot (dz = 0.626) at an alpha level of 0.05 with a power of $1-\beta = 0.95$. We increased the minimum sample to $N = 50$ to ensure we had a sample size suitable for an exploratory factor analysis (EFA) of language factors (de Winter et al., 2009).

Additionally, we preregistered a data collection stop rule based on a Bayesian significance test of the distributional learning effects obtained. For this Bayesian test, we used a prior of 0.17 (based on the distributional learning effect obtained in the pilot study) to determine if the distributional learning effect was substantially supportive of either the experimental hypothesis (i.e., that there would be an overall distributional learning effect; $BF > 3$), or the null hypothesis (i.e., that there would be no overall distributional learning effect; $BF < 0.33$) (Dienes, 2014).

2.3.1 Methods

2.3.1.1 Participants

56 participants (42 female, 14 male) were recruited from the student population of Nanyang Technological University in exchange for course credits or payment of $5 per half hour following completion of the study. Three male and three female participants were excluded from the study due to either a

failure to follow instructions, or for not being from our target demographic of Singapore Mandarin speakers. The remaining sample of 50 participants were aged between 18 – 25 years of age (*M* = 20.3, *SD* = 1.7). All participants self-identified as English-Mandarin bilinguals who had lived in Singapore for the majority of their lives. Informed consent was collected from all participants prior to commencement of the study, and this study was approved by the IRB of Nanyang Technological University institution (IRB-2019-01-034).

2.3.1.2 Equipment

The equipment was identical to that of the pilot study.

2.3.1.3 Stimuli

All visual and auditory stimuli were identical to that of the pilot study.

2.3.2 Procedure

The procedure for the distributional learning task and the language ID for the main study were identical to that of the pilot study.

2.3.3 Analysis plan

All analyses with the exception of Bayesian Factor analyses were conducted with R (R Core Team, 2020) Decision gradient slope values at pre-training and post-training test phase were obtained for each participant in the manner described in the pilot study. As the decision gradient slope values obtained for each participant in this study were not normally distributed, a log transform was carried out on all slope values to approximate a Gaussian distribution using the *log1p* function from the *base* package (R Core Team, 2020).

2.3.3.1 Pre-training control analysis

To ensure that exposure to the sound tokens presented prior to the bimodal distributional training phase had not led to any distributional learning effects, we first conducted a control t-test on the slope values of the first and third pre-training blocks to determine if significant changes in slope value had occurred prior to the training phase using the *t.test* function from the *stats* package (R Core Team, 2020).

2.3.3.2 Distributional learning analysis

Following this, we conducted our main analysis to determine if there were distributional learning effects following training, and to determine if Mandarin understanding scores were associated with distributional learning effects. In order to do this, we conducted a linear mixed effects models analysis in R (R Core Team, 2020) on our participants' individual log-transformed decision gradient slope values using the *lme* function from the *nlme* package (Pinheiro et al., 2022), with a categorical fixed factor of test phase (pre-training, post-training) and an interaction between test phase and self-reported Mandarin understanding scores, with random intercept of participant, and random by-participant slope for test phase. Bayes factors were calculated for each of the *p*-values obtained with a Bayes calculator created by Palif based on the 2008 Dienes Bayes calculator (accessible here:

https://bencepalfi.shinyapps.io/Dienes_BF_ calculator/). For this analysis, a preregistered prior of 0.17 was used for the categorial fixed factor of test phase. As an exploratory analysis, we used a prior of -0.29 for the interaction between test phase and Mandarin understanding. Both priors were obtained from the pilot study.

2.3.3.3 EFA of Learning Microstructure and Language Factors

To investigate if different aspects of language background were related to distributional learning effects at different points across the study, we conducted an EFA on different language and learning factors using the *princomp* function (R Core Team, 2020). For this EFA, we computed three language factors with scores derived from the Language ID. The three factors are as follows 1) Mandarin understanding score of each participant, 2) the total number of languages understood by each participant, and 3) the overall language understanding score of each participant. We also obtained three additional learning microstructure variables from the decision gradient slope values of individual test blocks in the distributional learning task. The three variables are as follows: 1) ‘no-training baseline’, 2) ‘plasticity’, and 3) ‘elasticity’ (see Figure 2.7).

2.3.3.3.1 Language factors

1) Mandarin understanding: The Mandarin understanding score of each participant was obtained directly from the number rating that each participant gave for Mandarin on the Language ID.

2) Total number of languages understood: The total number of languages understood for each participant was computed as the total number of languages the participant rated at 1 or higher on the language ID. This factor would allow us to determine if the experience of acquiring multiple languages might itself be related to any learning advantages.
3) Overall language understanding score: The overall language understanding score of each participant was computed as the total of the number ratings each participant gave on their Language ID. For instance, a participant who rated their English understanding at a 7, their Mandarin understanding at a 3, and their Malay understanding at a 5 would have an overall language understanding score of 15.

2.3.3.3.2 Learning microstructure variables

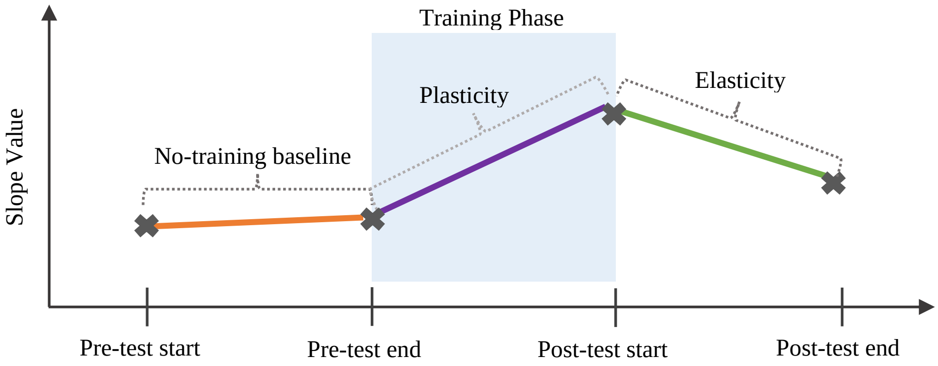
1) "No-training baseline": This variable was obtained to determine if any participants showed an increase in sensitivity to the Hindi dental-retroflex contrast continuum prior to training. The variable was obtained for each participant by deducting their slope value at the first pre-training block from their slope value at the last pre-training block. Positive values would indicate an improvement in perception prior to training.
2) "Plasticity": This variable was obtained to determine if participants showed a distributional learning effect immediately after training relative to their pre-training response patterns. This factor would allow us to examine "microlearning" changes within the duration of the post-

test phase and identify participants who showed large distributional learning effects immediately after training. This variable was obtained for each participant by deducting their slope value at the last pre-training block from the first post-training block. Positive values would indicate an improvement in perception immediately following training.

3) “Elasticity”: This variable was obtained to determine if learning attrition occurred over time following the end of the training phase. This factor would allow us to examine “microlearning” changes within the duration of the post-test phase and identify participants who may have shown distributional learning effects immediately after training, but then have rapidly lost this effect over time. This variable was obtained for each participant by deducting their slope value at the first post-training block from the slope value at the last post-training block. Negative values would indicate that learning attrition had occurred over time. Figure 2.7 depicts an example of a predicted learning microstructure trajectory across the study.

Figure 2.7

Depiction of learning microstructure across distributional learning task.

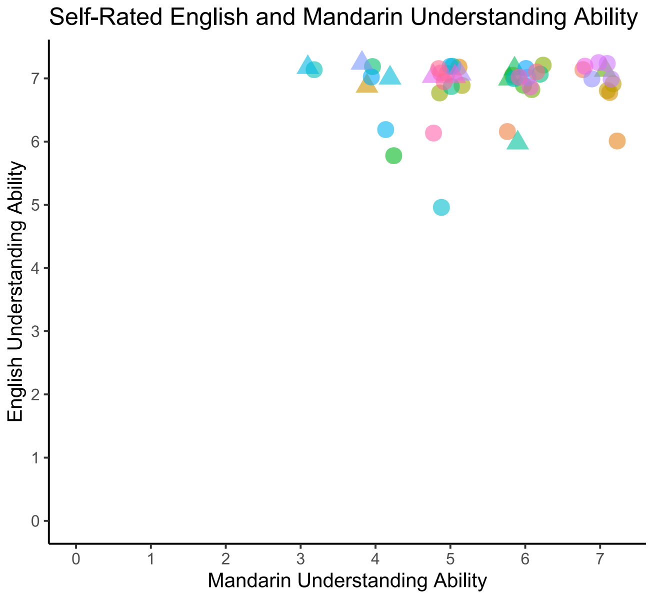


2.3.4 Results

2.3.4.1 Language ID

Figure 2.8

Distribution of participants' English and Mandarin proficiencies in main study.



Note. *N* = 50. Circles: female, triangles: male. Responses jittered by .25 for visualisation.

Figure 2.8 shows the distribution of our participants' self-rated English and Mandarin understanding abilities. Overall, our participants' self-reported English understanding scores were high (*Median* = 7, Range: 5-7). This indicated that the participants would have had not issues in understanding the experiment instructions given in English. On the other hand, their Mandarin understanding scores were more variable (Median = 6, Range: 3-7). No participants reported understanding Hindi (a non-Dravidian Indian language), but two participants reported understanding a few words of Tamil (a Dravidian Indian language), with both rating their Tamil understanding with a score of 1.

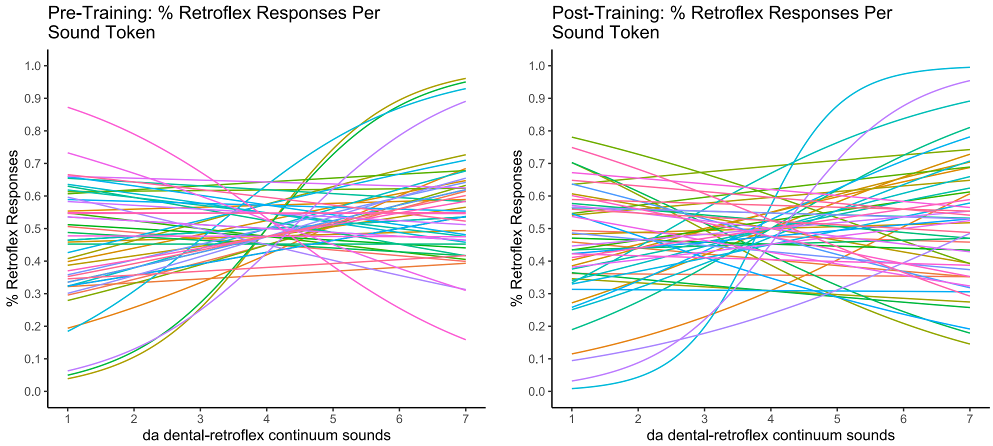
2.3.4.2 Pre-training control analysis

The control analysis revealed no significant difference in slope value between the first and third pre-training blocks, (*t*(49) = 1.04; *p* = .30; pre-1: *M* = .12, *SD* = .44; pre-3: *M* = .05, *SD* = .47). Thus, no significant learning had occurred prior to the bimodal distributional training phase.

2.3.4.3 Main distributional learning analysis

Figure 2.9

Decision gradient slopes for individual participants at pre- and post-training test phase.

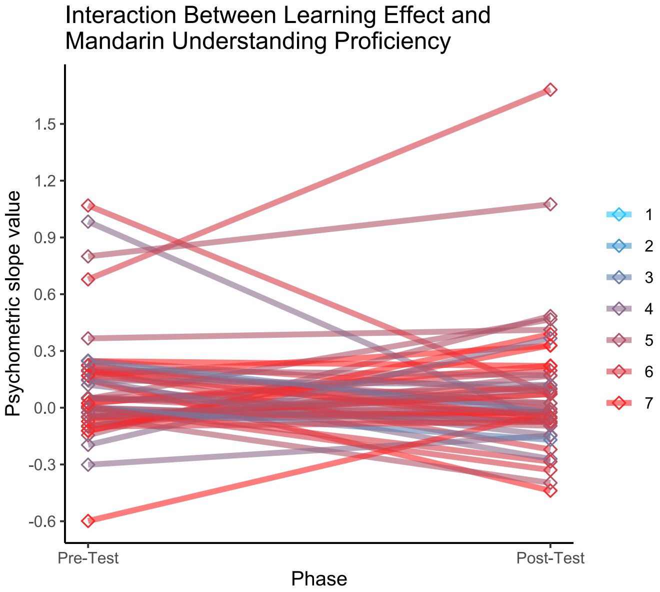


Note. *N* = 50. Pre-training test phase (left). Post-training test phase (right)

Fitted psychometric slopes for the pre-training and post-training test phases are presented in Figure 2.9. The linear mixed effects model analysis revealed no significant main effect of test phase (pre-training, post-training) on the slope values, and the corresponding Bayes factor obtained showed substantial evidence in support of the null hypothesis. This suggests that the data collected did not support an overall distributional learning effect following bimodal distributional training: $t(1,48) = -.54$; $p = .59$; $\eta p^2 = .005$; $B_{H(0, .17)} = .0001$; pre-test slope: $M = .06$, $SD = .25$; post-test slope: $M = 0.04$, $SD = .27$.

Figure 2.10

Visualisation of interaction between distributional learning effect and Mandarin proficiency.



Note. $N = 50$. Colours represent self-rated Mandarin understanding proficiency scores.

Figure 2.10 shows participants' pre- and post-training slope values coloured by each participant's self-reported Mandarin understanding score. The linear mixed effects model analysis did not reveal a significant interaction between test phase and self-reported Mandarin understanding scores. The corresponding Bayes factor obtained indicated that there was insufficient evidence from the data to support either the experimental hypothesis (i.e., the data supports an interaction effect between distributional learning effect and Mandarin understanding), or the null hypothesis (i.e., the data does not support an interaction effect between distributional learning effect and Mandarin understanding): ($t(1,48) = .31$; $p = .76$; $\eta p^2 = .002$; $B_{H(0, -.29)} = .72$).

2.3.4.3 Exploratory Factor Analysis of Learning Microstructure and Language Factors

2.3.4.3.1 Language factors

1) See *Language* ID section above for Mandarin understanding.
2) Participants reported understanding an average of 3.9 languages each (*SD* = 1.3, Range: 3-7).
3) Participants had an average overall understanding score of 17.1 (*SD* = 4.1, Range: 10-31).

2.3.4.3.2 Learning microstructure variables

1) Participants had an average ‘no-training baseline’ slope change of -.07 (*SD* =.48, range: -1.26 – 0.75)
2) Participants had an average ‘plasticity’ slope change of .04 (*SD* = .60, range: -1.65-1.33)
3) Participants had an average ‘elasticity’ change of 1.74 (SD = 14.0, range: -19.5-95.4)

2.3.4.3.3 Factor loadings

Table 2.1

Factor loadings on EFA of language factors and learning microstructure.

	Component					
	1	2	3	4	5	6
Eigenvalues	***2.41***	1.19	.92	.75	.65	.08
Percent Variance Explained	***40.2%****	19.8%	15.4%	12.5%	10.8%	1.31%
Factor loadings						
No-Training Baseline	.663	.130	.285	.092	.673	.019
Plasticity	-.474	.361	.507	-.623	.009	.009
Elasticity	.306	.687	.388	.428	-.316	-.010
Mandarin Understanding	-.419	.681	-.562	-.004	.182	.106
Total number of languages understood	-.802	-.309	.350	.329	.064	.171
Overall language understanding score	-.914	.095	.028	.246	.239	-.193

Note. Elasticity is inversely represented, with positive value indicating lower Elasticity. * Indicates principal components of interest that account for >20% of variance.

The results of the EFA conducted on these six language and learning factors revealed one component of interest – component 1 – that accounted for more than 20% of the total variance. Component 1 showed strong relationships between the language factors and plasticity. Individuals with a larger number of languages in their repertoire and a higher overall language understanding score were less likely to show changes in perception prior to training. At the same time, individuals with a higher Mandarin understanding score were also less likely to show changes in perception following training – however, this relationship was much weaker. See Table 1 for factor loadings on each component.

2.3.5 Discussion

Overall, the results of both the pilot study and the main study showed that *some* adults in the English-Mandarin bilingual population of Singapore may be sensitive enough to distributional cues to show distributional learning effects following bimodal distributional training on a novel Hindi /d̪a/-/ɖa/ dental-retroflex continuum. This can particularly be seen in the results of the small-scale pilot study which revealed a number of strong learners. However, the results of the well-powered main study do not support a significant overall distributional learning effect for the majority of our participants. It is likely that certain adult individuals may be more adept at extracting distributional cues from training stimuli – this was likely the case for the strong learners observed in the pilot study who are less representative of the general population of adult English-Mandarin bilinguals in Singapore. As the participants of both the pilot and main study were recruited from the same student population of NTU, it is unlikely that recruitment biases would have contributed to the different findings of each of the studies.

At the same time, the lack of an overall distributional learning effect within the participants of our main study was not entirely unexpected. Existing studies have shown that adults typically experience more difficulty in perceiving acoustic differences in novel, non-native, phoneme contrasts as compared to children and infants (e.g., Best & Strange, 1992), and are generally poorer at extracting distributional cues from training stimuli (e.g., Wanrooij et al., 2014). This is likely related to the manner in which language learning strategies shift from distributional learning with age, with adults typically favouring the use of more salient acoustic and lexical cues for the acquisition of novel speech sound contrasts (Barrios et al., 2022; Hayes-Harb, 2007; Liu, et al. 2022; Werker, 2018). As the study we adapted was originally conducted on Grade 3 Dutch children (Vandermosten et al., 2018), children in the original study were likely more sensitive to the differences in dental-retroflex ambiguity in the stimuli, and would have been more adept at extracting distributional cues from the training set in a much shorter period of time.

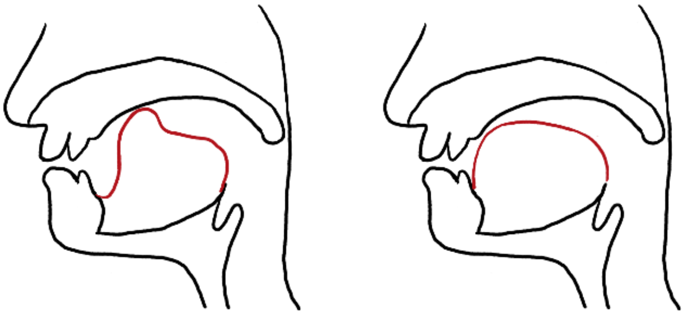
Additionally, while studies have shown that adults can show distributional learning effects from distributional learning paradigms, certain conditions typically have to be met within the nature of the training stimuli for successful learning to occur. For instance, studies that have successfully elicited distributional learning effects in adults typically utilise training stimuli with enhanced levels of acoustic constrast, akin to that of hyperarticulation present in infant-directed speech (e.g., Escudero et al., 2011), or utilise training stimuli with clear featural overlaps with the adult participants' native languages (e.g., Chládková & Šimáčková, 2021). The nature of the synthesised Hindi dental-retroflex /d̪a/-/ɖa/ stimuli used in our study could have compounded the difficulties faced by our adults in extracting distributional cues during training, as the stimuli may have been too finely tuned for successful adult-focused distributional learning.

As mentioned in the methods section of the pilot study, the dental and retroflex endpoints of the 7-step continuum were synthesised to be very close to the category boundary for these speech sounds in Hindi, making them harder to discriminate than in naturalistic Hindi speech (Vandermosten et al., 2018). Moreover, the degree of acoustic difference between each sound token was very small, with the tokens only differing slightly from each other on the third formant and central frequency of the initial burst (Golestani et al., 2002). This differs from naturalistic Hindi dental and retroflex speech sounds. In an acoustic analysis of naturalistic productions of the Hindi dental and retroflex stop phonemes, Verma and Chawla (2003) identified key differences in not only the third formant, but also the first, second, and fourth formants. These additional transition cues could be crucial for clear discrimination of the two phoneme categories, leading to additional ambiguity in the synthesised dental-retroflex contrast continuum used in this study. In fact, it is worth noting that one native Hindi speaking participant (excluded from data analysis) was unable to discriminate the sound tokens by ear, and did not demonstrate any distributional learning effect from the bimodal distributional training. It is therefore likely that the stimulus set used in this study would have fallen outside of the perceptual sensitivity window of most adults, making it unsuitable for eliciting distributional learning effects with a passive distributional learning task.

At the same time, it is also possible that there may have been an insufficient amount of featural overlap between retroflexion in Hindi and Mandarin for any transfer effects to occur. A feature of retroflex articulation has traditionally been applied to fricatives in Mandarin Chinese, and this can be particularly seen in formal Mandarin education materials which typically include instructions on producing retroflex frication with a raised and curled tongue tip similar to that of retroflexion in Hindi (e.g., Ou & Guo, 2014). However, it is possible that the actual lingual posture used to produce retroflexion in Mandarin might differ from the “curled” lingual posture typical to Hindi retroflex. An early x-ray study conducted by Ladefoged and Wu (1984) on three native standard Mandarin speakers from Beijing demonstrated that none of the speakers utilised a “curled” sub-apical tongue tip position for retroflexion, instead using alveolar constriction with the tongue body to produce “retroflex” frication. This was later described as a laminal post-alveolar place of articulation by Ladefoged and Maddieson (1996), in which the tongue tip is lowered, and post-alveolar constriction is made with the tongue blade curving upwards (also referred to as a “humped” retroflex in Luo, 2020). More recently, a lingual ultrasound study by Luo (2020) conducted on 18 native standard Mandarin speakers recruited in Yangzhou (a city in eastern China) found that the use of “curled” tongue tip retroflexion is extremely rare in mainland Mandarin speakers with only one example of true “curled” retroflexion found in 162 speech tokens. Instead, the “humped” place of articulation was found to be the most common lingual gesture used in the production of standard Mandarin retroflexion (Luo, 2020) – see Figure 2.11 for an example of the difference in lingual shape between the “curled” and “humped” retroflex places of articulation. While we are not currently aware of any studies that have examined the lingual gesture of retroflexion in the Singapore variety of Mandarin, it is possible that there may be a similar pattern of a “humped” rather than “curled” lingual gesture for retroflexion in the local variety of Mandarin. If this is indeed the case, we would not be able to assess shared feature transfer effects between Hindi and Mandarin retroflexion. At the same time, studies have shown that speakers of certain outer-circle varieties of Mandarin (including Singapore and Taiwan Mandarin) tend to exhibit deretroflexion, or phoneme contrast merger between the alveolar and retroflex places of articulation (Chang & Shih, 2015; Ng, 1985). Thus, even if Singapore Mandarin speakers do use a “curled” tongue tip gesture for retroflexion, there could still be a decreased amount of structural overlap between retroflexion in Hindi and Singapore Mandarin, making it difficult to accurately assess language-specific transfer effect in our adult participants with this set of Hindi dental-retroflex stimuli.

Figure 2.11

Lingual place of articulation for "curled" (left) and "humped" (right) retroflexion.



The lack of an overall distributional learning effect in our main study also likely contributed to the lack of evidence of an interaction effect between distributional learning effects and our participants' self-reported Mandarin understanding abilities. At the same time, the Language ID that we used to measure self-reported Mandarin understanding abilities may not have been able to capture a sufficiently meaningful amount of individual variance on a scale of 0 to 7. While a traditional test of language proficiency would likely provide us with a more detailed measure of Mandarin understanding ability in our participants, a validated, empirical measure for examining Mandarin proficiency in adult speakers of the Singapore variety of Mandarin does not exist at this point in time. Thus, the use of a self-rated language understanding scale allows us to obtain important information on the relative differences in our participants' Mandarin understanding abilities that we would otherwise not be able to access. Moreover, studies on second language acquisition have shown that self-assessment of language abilities can provide an accurate measure of linguistic ability in a homogenous group of bilinguals, with self-rated proficiencies being correlated with empirical language proficiency tests (e.g., Thompson, 2015). Moving forward, we will continue our use of the self-rated Language ID, while increasing the range of the scale to 0 to 100.

Our use of a power analysis to determine the minimum sample size of our main study ensured that our study was well-powered enough to obtain a distributional learning effect if one did indeed exist, and our use of Bayesian hypothesis testing enabled us to better qualify the null results that we found. While conventional *p*-values can only tell us the likelihood of our results being due to a false positive, the Bayes factor gives us a weighted estimation of how supportive our data are of either the experimental hypothesis, or the null hypothesis (Dienes, 2016; Lakens et al., 2018). The Bayes factor we obtained for our analysis of a distributional learning effect in the main study thus showed that the data we collected were around 1000 times more supportive of the null hypothesis rather than the experimental hypothesis (i.e., that phoneme contrast sensitivity would increase at post-training as compared to pre-training). Thus, we are confident that the paradigm and stimuli of this study were not suitable for eliciting distributional learning effects in a well-powered sample size of Singapore English-Mandarin bilingual adults. This demonstrates the value of selecting a minimum sample size based on a power analysis, and using a

Bayesian inference threshold to ensure that sufficient data is collected to obtain meaningful results in a study.

2.4 Conclusion

On the whole, studies have suggested that language learning advantages in bilinguals may be linked to language-specific transfer effects arising from featural overlaps between known and to-be-learnt languages. We attempted to explore this hypothesis in adult English-Mandarin bilinguals in Singapore with a bimodal distributional learning paradigm. While we found that some adult participants were able to show strong distributional learning effects for a synthesised Hindi dental-retroflex /d̪a/-/ɖa/ contrast following bimodal distributional training, these results of these participants were not generalisable to the larger population. As a result of this, we were unable to obtain evidence of an overall distributional learning effect or of a language-specific transfer effect in our large-scale preregistered main study.

Existing studies have established that sensitivity for novel phoneme contrasts as well as the ability to utilise distributional cues for learning tend to decline with age. Thus, it is likely that this bimodal distributional learning paradigm paired with an acoustically fine-grained stimulus set would be better suited to younger participants (e.g., Vandermosten et al., 2018) rather than the adult participants in our target population. At the same time, research published after this study was completed suggest that Mandarin and Hindi may not share as much overlap in retroflexion as we previously thought (Luo, 2020; Tiede et al., 2019). This could make the stimuli used in this study unsuitable for assessing language-specific transfer effects within Singapore English-Mandarin bilingual adults.

We believe that a distributional learning paradigm is still appropriate for our investigation as, studies have shown that adults *can* show strong distributional learning effects, particularly when there are clear featural overlaps between known languages and to-be-learnt speech sound stimuli (e.g., Chládková & Šimáčková, 2021; Chládková et al., 2022). In order to accurately assess distributional learning and language-specific transfer effects in Singapore English-Mandarin bilinguals, we would first have to obtain a clearer understanding of the phonology of the under-documented contemporary Singapore variety of Mandarin. This would allow us to identify and create a more appropriate set of training stimuli aligned to the acoustic norms of speech sounds in contemporary Singapore Mandarin.

As very little documentation has been conducted on the nature of the alveolar-retroflex contrast in the Singapore variety of Mandarin, we thus commenced a series of novel investigations into the articulatory and acoustic characteristics of the alveolar-retroflex contrast in Singapore Mandarin. We first set out to confirm our hypothesis that Hindi and Singapore Mandarin retroflexion do not, in fact, share any similarities in tongue posture. In the next chapter we thus conducted a lingual ultrasound imaging study to investigate the lingual gesture used for “retroflex” phonemes in Singapore Mandarin.

Chapter 3: Investigating the Singapore Mandarin alveolar-retroflex sibilant place of articulation with lingual ultrasound imaging

3.1 Introduction

Standard Mandarin ostensibly features a phoneme contrast between alveolar and retroflex places of articulation for sibilants. In educational material for formal instruction of Mandarin, the "retroflex" place of articulation is typically described as a "curled" retroflex characterised by the tongue tip curling upwards and towards the palatal region (e.g., as described and depicted in Ou & Guo, 2014), similar to that of retroflexion in Hindi stop phonemes. However, studies examining the lingual shape of retroflex phonemes produced by Mandarin speakers suggest that there may in fact be variation in the tongue shape in which "retroflex" frication is produced in Mandarin. An early x-ray study by Ladefoged and Wu (1984) found that out of three Beijing Mandarin speakers, none utilised true "curled" retroflexion in their production of the Mandarin phoneme /ʂ/, instead utilising a flat laminal post-alveolar place of articulation to produce this type of frication in speech (Ladefoged & Maddieson, 1996). A more recent lingual ultrasound study by Luo (2020) found that while true "curled" retroflexion does occur in mainland Mandarin speech, it is extremely rare, with only one instance of "curled" retroflexion observed out of 162 speech tokens recorded from 18 native standard Mandarin speakers. Instead, Luo (2020) found that contemporary standard Mandarin speakers are more likely to produce retroflexion with a "humped" tongue position similar to that of the descriptions of Ladefoged and Wu (1984) and Ladefoged and

Maddieson (1996) (see Figure 2.11 for example of “humped” retroflexion), or less frequently with a “bunched” position characterised by constriction of the tongue tip in the alveolar region, and tongue root retraction. From these studies, we can see that the “retroflex” tongue position used by native speakers of the standard mainland variety of Mandarin typically differs significantly from classical “curled” retroflexion as found in Hindi and other Dravidian Indian languages. It is possible that this may be the case for Singapore Mandarin speakers as well. However, the results of the studies discussed above cannot be generalised to speakers of the outer-circle variety of Singapore Mandarin. While there is very little documentation on the phonology of Singapore Mandarin, early observational studies in the 1980s suggest that Singapore Mandarin speakers do not produce Mandarin alveolar-retroflex contrast frication in the same manner as that of standard Mandarin speakers. Instead, there is likely to be a degree of alveolar-retroflex contrast merger, or deretroflexion, in Singapore Mandarin due to heavy influence of Southern Min languages which do not feature a contrast between alveolar sibilants and a post alveolar/retroflex place of articulation (Lock, 1989, Ng, 1985). Thus, is it likely that contemporary speakers of Singapore Mandarin may exhibit differences in production of the alveolar-retroflex contrast (particularly in production of retroflexion) in comparison to speakers of the standard variety of Mandarin.

Similar to the Singapore variety of Mandarin, Taiwan Mandarin is also heavily influenced by non-retroflex Southern Min languages and features different degrees of deretroflexion between speakers (Chang & Shih, 2015;

Kubler, 1985). The effect of this influence can be seen in recent lingual ultrasound studies conducted on alveolar-retroflex frication produced by Taiwan Mandarin speakers. Chiu et al. (2020) investigated Mandarin frication production in seven native Taiwan Mandarin speakers and found that two participants consistently showed no significant difference in tongue posture for alveolar and retroflex frication – typical of total phoneme contrast merger. On the other, hand two participants consistently showed significant differences in tongue alveolar and retroflex frication, and the remaining three showed varying degrees of overlap and non-overlap in different vowel contexts (Chiu et al., 2020). An electromagnetic articulometry (EMA) and lingual ultrasound study by Tiede et al. (2019) conducted on three native Taiwan Mandarin speaking participants found that while all three participants produced Mandarin alveolar and retroflex frication contrastively, the difference in tongue posture for each place of articulation was small and largely found in slight differences in location of tongue-tip constriction (further back in the alveopalatal region for retroflexion). At the same time, the study by Tiede et al. (2019) found consistent upward curling of the tongue tip for alveolar constriction in retroflex frication (albeit without sub-apical contact with the alveopalatal region as in Hindi retroflexion). This is different from the typical “humped” retroflex tongue position produced by standard Mandarin speakers, in which there is no evidence of tongue-tip alveolar constriction (Luo, 2020). Thus, these studies on lingual imaging of Taiwan Mandarin show that the production of alveolar-retroflex frication differs between varieties of Mandarin. In particular, speakers of varieties of Mandarin that are influenced by non-retroflex languages are likely to show different degrees of contrast merger between alveolar and retroflex places of articulation. The typical tongue posture used for the production of “retroflexion” may also differ between different varieties of Mandarin. In order to investigate the lingual posture used for the production of alveolar-retroflex frication in contemporary Singapore Mandarin, we conducted a preliminary observational lingual ultrasound study on five Singapore English-Mandarin bilingual adults. This study would allow us to confirm our hypothesis that contemporary Singapore Mandarin speakers do not use a “curled” lingual place of articulation in retroflexion, and hence, cannot transfer their mental representation for this place of articulation in a Hindi phoneme learning task.

3.2 Methods

3.2.1 Participants

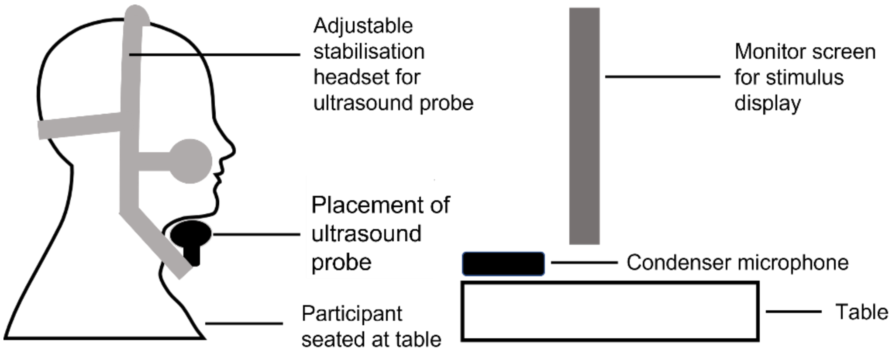
Based on the sample sizes of existing lingual imaging studies conducted on Mandarin speakers, we focused this study on data collected from five participants (3 female, 2 male). All participants were recruited from the student population of Nanyang Technological University in exchange for course credits or payment of $10 following completion of the study. The participants were aged between 21 – 28 years of age (*Median* = 24.0, *SD* = 2.6), all of whom reported spending their entire lives in Singapore. Informed consent was collected from all participants prior to commencement of the study, and this study was approved by the IRB of Nanyang Technological University institution (IRB-2018-09-002).

3.2.2 Equipment

The speech production task was presented to participants on a 10-inch LCD screen on a Microsoft Surface Go computer. Audio recordings of participants’ speech tokens were captured with an AT3035 Audio-Technica cardioid condenser microphone powered by a Focusrite Scarlett Solo audio interface. All recordings were made in stereo WAV format, with a sample rate of 44.1kHz and a bit depth of 16 bits. Ultrasound video recordings of the sagittal plane of the tongue were captured with an MC10-5R10S-3 element (10mm, 5-10MHz) microconvex ultrasound probe using the SonoSpeech micro ultrasound system operated with Articulate Assistant Advanced version 2.17.06 (AAA; Articulate Instruments), which synchronised all audio and video signals. A lightweight aluminium stabilisation headset was used to secure the ultrasound probe in place under the participants’ chin during data collection. Figure 3.1 contains a schematic of the setup used for data collection.

Figure 3.1

Ultrasound setup for participant data collection.



Note. Not drawn to scale.

3.2.3 Stimuli

Visual stimuli consisted of 12 Mandarin alveolar-retroflex words from 6 minimal pairs. The words were chosen following consultation with five fluent local Singapore Mandarin speakers to ensure that this task would be well suited for participants with different proficiencies in Mandarin. Table 3.1 contains the list of words. On each trial, participants were presented with the Chinese character of each word and its corresponding pronunciation in Hanyu Pinyin. Written stimuli were also included to remind the participant of the language that the word was in. Figure 3.2 depicts an example of the visual stimuli participants saw on one trial. Participants also read 25 English words for a separate study. The recordings for these words were not analysed for this study (see Appendix 3.1). No audio stimuli were used for this speech production task.

Table 3.1

List of Mandarin alveolar-retroflex words in speech production task.

No.	Minimal Pair	IPA	Place of articulation	Character/Pinyin	Word Meaning
1	1	[sǎi]	Alveolar	赛/sài	competition
2		[ʂǎi]	Retroflex	晒/shài	sun shining upon
3	2	[sán]	Alveolar	三/sān	three
4		[ʂán]	Retroflex	山/shān	mountain
5	3	[sə́n]	Alveolar	森/sēn	many trees
6		[ʂə́n]	Retroflex	身/shēn	body
7	4	[tsə́ŋ]	Alveolar	增/zēng	increase
8		[ʈʂə́ŋ]	Retroflex	筝/zhēng	kite
9	5	[tsʰǒu]	Alveolar	凑/còu	gather together
10		[ʈʂʰǒu]	Retroflex	臭/chòu	smelly
11	6	[tsʰuə́n]	Alveolar	村/cūn	village

12	[ʈʂʰuə́n]	Retroflex	春/chūn	spring (season)

Figure 3.2

Example of visual stimuli presented on a trial of the word reading task.

Mandarin

村/cūn

Note. Example stimuli depicts the word [tsʰuə́n]. Not drawn to scale.

3.3 Procedure

3.3.1 Ultrasound speech production task

Participants were first introduced to the speech production task with written instructions in English. Participants were informed that a word would appear on the screen, and their task was to read each word aloud three times after being verbally prompted to by the researcher in spoken English. To ensure that speech was produced as naturalistically as possible, participants were instructed to pronounce the words in a manner similar to that of daily speech. Emphasis was placed on the fact that the task was not a test of pronunciation accuracy, and that we were simply interested in documenting the unique speech patterns of Singapore Mandarin speakers.

Following the instructions, participants completed one block of audio recording consisting of 12 trials. For each trial, participants were presented with one of the 12 Mandarin alveolar-retroflex words, and read each word three times. Figure 3.2 shows an example of the visual stimuli presented in one trial. After reading each word three times, the participant made a keypress to move on to the next trial. The 12 Mandarin words were presented to each participant in random order. No feedback was given to the participants so as to not influence their speech production. No trial timeout was implemented to ensure that participants had sufficient time to read all the words presented.

3.3.2 Language ID

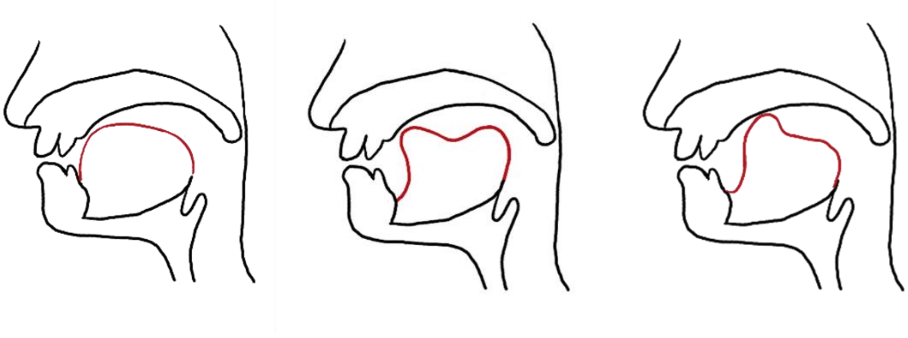
To collect data on participants' language understanding abilities, they were asked to rate how well they understood the four main languages in Singapore (English, Mandarin, Malay, and Tamil) on a scale of 0 to 100 using an onscreen sliding scale. Participants were also given an option to add in more languages or dialects to rate.

3.4 Analysis Plan

As this was a preliminary observational study, visual inspection was conducted for each ultrasound recording made for each retroflex sound token. The frication portion of each recorded speech token was identified from the waveform window in AAA, and the peak lingual position of the frication was identified by eye. Lingual positions for production of Mandarin retroflex phonemes were classified into three main lingual shapes based on the descriptions of Luo (2020) – “humped”, “bunched”, and “curled”. The humped lingual position of Mandarin retroflexion is characterised by a single, highly peaked point of laminal post-alveolar constriction (Luo, 2020). The bunched lingual position of Mandarin retroflexion is characterised by two points of constriction with anterior alveolar constriction, and tongue root retraction (Luo, 2020). The curled lingual position of Mandarin retroflexion is akin to “true” retroflexion found in Hindi retroflex phonemes and is characterised by the curling of the tongue tip upwards and backwards towards the palatal region (Luo, 2020). Figure 3.3 shows vocal tract diagrams of the humped, bunched, and curled lingual positions described in Luo (2020). The number of incidences of each retroflex lingual posture observed were recorded for all five participants.

Figure 3.3

Lingual posture for "humped" (left), "bunched" (middle), and "curled" (right) retroflexion.



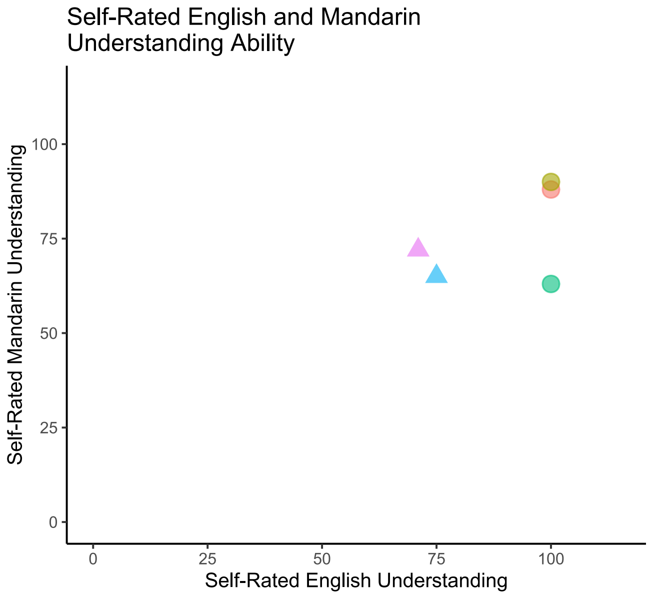
Visual inspection was also conducted for the lingual positions of the Mandarin alveolar phonemes, and the number of incidences of each alveolar lingual position were recorded for all five participants. Based on Luo (2020) and Tiede et al. (2019), we do not expect to find observably different lingual shapes used for alveolar as compared to retroflex frication. Differences between the two places of articulation are more likely to lie in slight differences in constriction location – further back for alveolars in Mainland Mandarin, and further forward for alveolars in Taiwan Mandarin (Luo, 2020; Tiede et al., 2019).

3.5 Results

3.5.1 Language ID

Figure 3.4

Participants' self-reported English and Mandarin understanding ability.



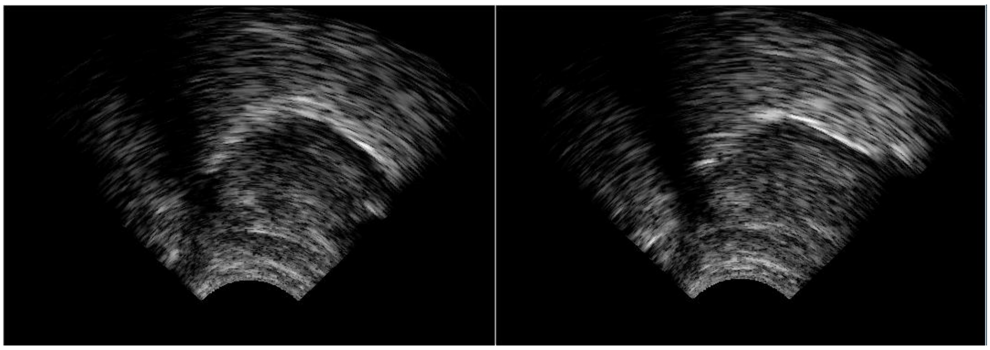
Note. *N* = 5. Circles: female, triangles: male.

Overall, participants self-reported high levels of understanding abilities for spoken English (Range: 71–100, *M* = 89.2, *SD* = 14.9), indicating that they would not have had any issues in understanding the task instructions. Participants self-reported more varying levels of understanding abilities for spoken Mandarin (Range: 63–90, *M* = 75.6, *SD* = 12.7). Figure 3.4 shows the distribution of participants' self-rated English and Mandarin understanding abilities.

3.5.2 Ultrasound speech production task

Figure 3.5

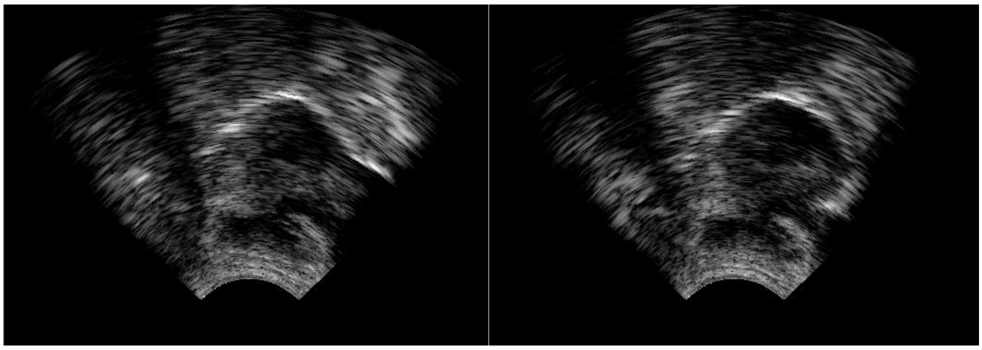
Raw lingual ultrasound images of "humped" tongue shape used in production of alveolar [sə́n] (left) and retroflex [ʂə́n] (right) frication.



Note. Recorded from Participant 3, female.

Figure 3.6

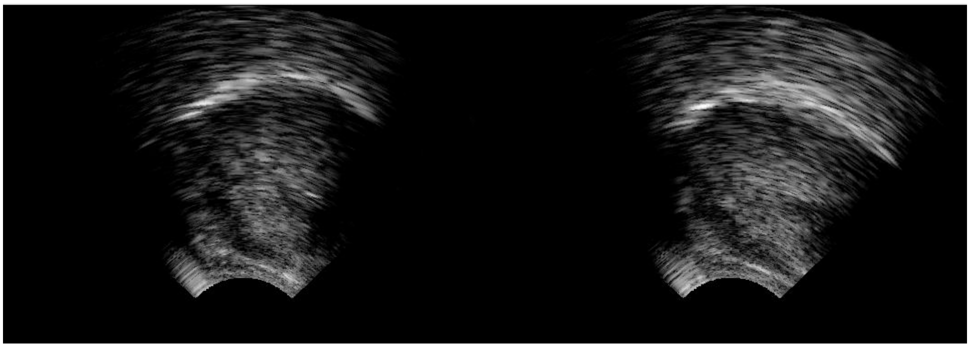
Raw lingual ultrasound images of "humped" tongue shape used in production of alveolar [tsʰuə́n] (left) and retroflex [ʈʂʰuə́n] (right) frication.



Note. Recorded from Participant 5, male.

Figure 3.7

Raw lingual ultrasound images of "humped" tongue shape used in production of alveolar [tsə́ŋ] (left) and retroflex [ʈʂə́ŋ] (right) frication.



Note. Recorded from Participant 1, female.

Table 3.2

Number of observations of each Mandarin retroflex lingual position.

Lingual position for Mandarin retroflex frication	Number of observations
Humped	90
Bunched	0
Curled	0

Note. Based on visual inspection of 90 ultrasound imaging recordings.

Table 3.3

Number of observations of each Mandarin alveolar lingual position.

Lingual position for Mandarin alveolar frication	Number of observations
Humped	90
Bunched	0
Curled	0

Note. Based on visual inspection of 90 ultrasound imaging recordings.

Figure 3.5 shows examples of the humped lingual positions of alveolar [sə́n] and retroflex [ʂə́n] frication produced by participant 3. Figure 3.6 shows examples of the humped lingual positions of alveolar [tsʰuə́n] and retroflex [ʈʂʰuə́n] frication produced by participant 5. Figure 3.7 shows examples of the humped lingual positions of alveolar [tsə́ŋ] and retroflex [ʈʂə́ŋ] frication produced by participant 1. All ultrasound images reported for this study are represented with the dental region of the oral cavity on the right (near the tongue tip), and the pharyngeal region on the left (near the tongue root).

Table 3.2 reports the number of incidences of each lingual position of retroflex frication that were observed for all speech tokens from all five participants. Table 3.3 reports the number of incidences of each lingual position of alveolar frication that were observed for all speech tokens from all five participants.

3.6 Discussion

Our observational study confirmed the predictions arising from Chapter 2. None of our five participants showed any evidence of a “curled” lingual tongue position in their productions of Singapore Mandarin retroflex frication. This is consistent with earlier studies on other varieties of Mandarin which show that it is extremely rare for Mandarin speakers to utilise “curled” retroflexion in speech. We also did not find any evidence of a “bunched” tongue position within the frication productions of our participants. Based on the findings of Luo (2020), the “bunched” tongue position for retroflexion is uncommon amongst Mainland standard Mandarin speakers, with only nine instanced of “bunched” retroflexion being observed in their study (out of 162). All five of our participants appear to use a “humped” tongue position for the production of Singapore Mandarin retroflex frication, wherein post-alveolar constriction is made with the tongue body. This appears to be similar to the flat laminal post-alveolar place of articulation described by (Ladefoged & Maddieson, 1996). There also appear to be noticeable individual differences in the degree of lingual curvature produced by each of our participants. Future studies can investigate these individual differences in greater detail.

At the same time, the ultrasound recordings of our participants appear to show that Singapore Mandarin speakers use a “humped” tongue shape for the production of both alveolar and retroflex frication, with different locations of peak constriction for each place of articulation. This is consistent with studies by Luo (2020) and Tiede et al. (2019) which have found that both standard Mandarin and Taiwan Mandarin speakers show little difference in overall tongue shape in their production of alveolar and retroflex frication, with differences arising mainly in tongue constriction location. Based on our observations of our Singapore Mandarin speaking participants, the location of peak constriction for alveolar frication appears to be more posterior in the oral cavity as compared to retroflex frication. This pattern appears to be aligned with that of standard Mandarin speakers (alveolar frication constriction more posterior than retroflex; Luo, 2020), as compared to Taiwan Mandarin speakers (alveolar frication constriction more anterior than retroflex; Tiede et al., 2019). It is also unclear at this stage whether the Singapore Mandarin alveolar-retroflex phoneme pair /s/ and /ʂ/ may overlap with the Singapore English alveolar-post-alveolar phoneme pair /s/ and /ʃ/ as there is no current existing acoustic documentation on frication contrasts in contemporary Singapore English. It is worth noting, however, that while the Singapore Mandarin alveolar-retroflex fricative contrast /s/-/ʂ/ may have some overlaps in acoustic space with the Singapore English alveolar-post-alveolar fricative contrast /s/-/ʃ/, the alveolar-retroflex affricate contrast /tsʰ/-/ʈʂʰ/ is only found in Mandarin but not English.

We plan to explore these points of interest in a preregistered follow-up study. This follow-up study will utilise General Additive Mixed Modelling (GAMM) and Principal Component Analyses (PCA) as conducted by Lin and Moisik (2019) on manual traces of peak tongue contours of alveolar and retroflex frication to identify statistically significant differences between alveolar and retroflex places of articulation in contemporary Singapore Mandarin and contemporary Singapore English. Data collection for this study has been completed.

Overall, we can infer from this study that Hindi and Singapore Mandarin do not have an overlap in terms of the place contrasts of the dental-retroflex contrast in Hindi, and the alveolar-retroflex contrast in Singapore Mandarin. We now have a clearer understanding of the articulatory characteristics of “retroflex” frication in Singapore Mandarin, which does not include tongue tip curling as typically found in retroflexion in other languages. For the purposes of this project, however, we will continue to refer to this phoneme contrast of interest as the ‘alveolar-retroflex’ phoneme contrast to maintain consistency with the nomenclature of existing literature on this Mandarin frication contrast. We remain primarily interested in the acoustic nature of this contrast, and the manner in which Singapore English-Mandarin bilinguals perceive and produce it, regardless of the naming conventions used to refer to this contrast. The fact that we were unable to obtain an overall distributional learning effect in the main study of Chapter 2 is likely related to the fact that the Singapore Mandarin “retroflex” does not share the typical apical tongue tip retroflexion of Hindi. While some studies have shown that adults *can* successfully show distributional learning effects following bimodal distributional training, these studies typically utilise training stimuli that share clear featural overlaps with the languages in the participants’ language repertoires (e.g., Chládková & Šimáčková, 2021). The lack of this expected overlap in Chapter 2 likely made it harder for our adult participants to extract distributional cues from the training stimuli. At the same time, this lack of overlap also means that we would not be able to accurately assess for language-specific transfer effects in Singapore English-Mandarin bilingual adults by training them on the Hindi dental-retroflex stimuli used in Chapter 2.

3.7 Conclusion

This observational lingual imaging study demonstrates that the lingual position of retroflexion present in Singapore Mandarin is indeed different from the “curled” tongue position typically seen in Hindi stop phonemes. Instead, Singapore Mandarin speakers appear to use a flat laminal post-alveolar place of articulation in production of retroflexed frication (as in Ladefoged and Maddieson, 1996). This shows that the Hindi dental-retroflex contrast stimuli we used in our distributional learning study in Chapter 2 were likely unsuitable for our investigation of a language-specific transfer effect in our target population of Singapore English-Mandarin bilinguals. In order to continue our project, we need to obtain a more suitable set of stimuli that is valid to the characteristics of the variety of Mandarin spoken by contemporary Singapore adults.

At the same time, this study is the first time in which lingual imaging has been conducted to investigate the phonology of Singapore Mandarin. Based on the results of this study, we plan to conduct a preregistered follow-up study on a larger, more representative sample size to determine if the results we have found are generalisable to other Singapore Mandarin speakers. We also plan to investigate the nature of the lingual place contrast between alveolar and retroflex frication in Singapore Mandarin in greater detail.

Chapter 4: Documentation of the perception and production of the contemporary Singapore Mandarin alveolar-retroflex sibilant contrast

4.1 Introduction

4.1.1 Variations in Mandarin alveolar-retroflex contrast

From the results of our lingual ultrasound study, we can see that retroflexion in Singapore Mandarin utilises a different place of articulation from “curled” retroflexion as found in Hindi stop phonemes. Thus, the Hindi dental-retroflex /d̪a/-/ɖa/ contrast stimuli used in our first distributional learning study was very likely unsuitable for examining language-specific transferadvantage in Singapore English-Mandarin bilingual adults. In order to accurately examine this learning advantage in our target population, we would have to use training stimuli that is more suited to the phonological norms of Singapore Mandarin speakers. While stimuli drawn from existing Mandarin corpora such as The Chinese Lexicon Project (with mainland Mandarin speakers, Sze et al., 2014) and the two-character compound word Chinese Lexicon Project (with Hong Kong Mandarin speakers, Tse et al., 2017) would be a *closer* match to the linguistic repertoire of our participants, phonological variations in different varieties of Mandarin mean that such stimuli would not be the *best* match to the variety of Mandarin spoken in Singapore. In particular, while the lingual ultrasound study showed that the majority our Singapore Mandarin speaking participants produced the retroflex phoneme with a “humped” lingual gesture similar to that of standard Mandarin speakers (Ladefoged & Maddieson, 1996; Luo, 2020), there still remains a degree of ambiguity in how similar the acoustic nature of the alveolar-retroflex contrast is in Singapore Mandarin as compared to standard Mandarin.

As discussed earlier, the nature of the alveolar-retroflex contrast varies in different varieties of Mandarin. Studies have shown that there is generally a clear alveolar-retroflex phoneme contrast in the standard variety of Mandarin (Chang, 2012). However, documentations of outer-circle varieties of Mandarin such as Taiwan Mandarin show that heavy influence from non-retroflex Southern Min Chinese languages tend to result in different degrees of deretroflexion in speech, wherein speech productions of the Mandarin retroflex sibilants /ʂ/, /ʈʂ/, and /ʈʂʰ/ tend to be more alveolar in nature, or are sometimes entirely replaced with the alveolar phonemes /s/, /ʈs/, and /ʈsʰ/ (Chuang & Fon, 2010; Chang, 2012; Tiede et al., 2019).

4.1.2 Singapore Mandarin alveolar-retroflex contrast

Singapore Mandarin is also an outer-circle variety of Mandarin that is heavily influenced by non-retroflex Chinese languages. As the majority of early Chinese migrants to Singapore arrived through seafaring migration from the coastal South-Eastern regions of China, the Chinese languages that were initially brought to Singapore consisted of non-retroflex Chinese languages from those regions – in particular, Hokkien and Teochew. The influence of these non-retroflex varieties can be seen in observational studies of Singapore Mandarin speakers conducted in the 1980s which suggest that there is an almost complete merger of the alveolar and retroflex categories of phonemes in

Singapore Mandarin, with the retroflex usually being replaced by the alveolar (e.g., Lock, 1989; Ng, 1985). However, while these early studies provide valuable insight into the nature of the Singapore Mandarin alveolar-retroflex contrast in the 1980’s, there is little information on the structural and acoustic nature of this contrast in contemporary Singapore Mandarin. In fact, more recent studies have shown that there may be a shift the degree of deretroflexion present in the variety of Mandarin spoken in present day Singapore due to the standard variety of Beijing Mandarin being increasingly perceived as a desirable prestige variety of Mandarin amongst younger Singaporeans (Chong & Tan, 2013).

Therefore, if we are to accurately assess language-specific transfer effects using the alveolar-retroflex contrast in Singapore English-Mandarin bilingual adults, we would have to account for these variations in the phonology of Singapore Mandarin. We therefore decided to synthesise our own distributional training stimuli based on the acoustic norms of the alveolar-retroflex contrast in the local variety of Singapore Mandarin. In order to accomplish this, we first needed to conduct a detailed acoustic analysis of this phoneme contrast in Singapore Mandarin as documentation on the acoustic nature of contemporary Singapore Mandarin does not yet exist in any current literature.

In this chapter, we therefore set out to closely examine the perception and production of the Singapore Mandarin alveolar-retroflex contrast in

English-Mandarin bilingual young adults. To do this, we designed and conducted two novel tasks. One to broadly document the accuracy of alveolar-retroflex speech perception with a word identification task, and the second to build a corpus based on the Singapore Mandarin alveolar-retroflex contrast with a spoken word production task. Using this data, we aimed to determine if word identification accuracy could be used as a direct measure of Mandarin alveolar-retroflex contrast familiarity. Crucially, we also planned to conduct detailed acoustic analyses on speech productions of the alveolar-retroflex contrast in Singapore Mandarin to obtain a clearer understanding of the acoustic nature of this contrast.

To our knowledge, this is the first large-scale documentation and analysis of the acoustic properties of Singapore Mandarin in existing literature. Our large-scale corpus of Singapore-Mandarin alveolar-retroflex words produced by 44 different Singapore Mandarin speakers is open source and is available online in the Spring-Village Corpus of minimal pairs in Singapore Mandarin Chinese (https://doi.org/10.21979/N9/ZTMPML). We have also chosen to make all the visual and audio stimuli of our novel tasks available in a central repository (https://doi.org/10.21979/N9/PBUD6A) to contribute to open science practices and facilitate extensions and replications of our work on Singapore Mandarin.

4.1.3 Documentation of alveolar-retroflex contrast in Singapore Mandarin

4.1.3.1 Perception of alveolar-retroflex contrast in Singapore Mandarin

To explore our participants' perception of the alveolar-retroflex contrast in Singapore Mandarin, we created a novel word identification task (the Spring-Village word identification task) with 24 picture cards depicting 12 minimal pairs of Mandarin alveolar and retroflex words. For this task, participants were presented with matching minimal pairs of picture cards and were instructed to identify the picture card that matched an auditory target word (presented in either Singapore or Beijing Mandarin). We designed this task to look for individual differences in how accurate Singapore Mandarin speakers are at reliably distinguishing between alveolar and retroflex-initial words in both Singapore as well as Beijing Mandarin. While this has yet to be explored in Singapore Mandarin speakers, a study conducted by Shih and Kong (2011) found that some Taiwan Mandarin speakers may experience a degree of confusion when tasked with identifying alveolar and retroflex words in Beijing Mandarin. Therefore, it is possible that we will also find similar patterns of responses of word identification confusion amongst our target population of Singapore English-Mandarin bilingual adults.

4.1.3.2 Production of alveolar-retroflex contrast in Singapore Mandarin

To collect data on our participants' speech productions of the alveolar-retroflex contrast in Singapore Mandarin, we created a novel word reading task (the Spring-Village speech production task) in which participants were presented with one of the 24 picture cards at a time, and were recorded reading each word out loud three times each. The speech production data obtained from this task was used in our fine-grained acoustic analysis of the Singapore Mandarin alveolar-retroflex contrast. Some studies have suggested that speech tokens obtained from word reading tasks may contain more formal articulations of speech (and potentially more retroflexion) as compared to speech tokens extracted from naturalistic casual conversation (e.g., Jeng, 2006). However, the collection and extraction of a large corpus of alveolar and retroflex speech tokens from casual Singapore Mandarin conversation was not feasible at this point in time, as the main researcher conducting these studies is not a speaker of any variety of Mandarin. In order to ensure that the speech tokens recorded were as naturalistic as possible, strong emphasis was placed on the fact that this task is not a test of reading accuracy, but a documentation of unique local speech sounds. Participants were thus encouraged to speak each of the words as they would in everyday life.

To examine the acoustic nature of the speech tokens we obtained, we chose to conduct two fine-grained acoustic analyses of the alveolar-retroflex contrast. For our first acoustic analysis of the speech production data, we focused on a centre of gravity (CoG) measurement of the frication portion of each word to examine overall differences between the alveolar and retroflex categories of speech sounds in Singapore Mandarin. The CoG is calculated from the weighted average of frequencies across a frication spectrum. The use of CoG values is widely considered to be a robust method of quantifying acoustic differences between alveolar and retroflex frication both within and between different varieties of Mandarin such as Beijing and Taiwan Mandarin (Chang, 2012; Chuang & Fon, 2010; Chiu et al., 2020; Tso, 2017). Existing studies examining Mandarin frication CoG values have demonstrated that alveolar frication tend to generate higher CoG values due to a greater degree of lingual constriction near the front oral cavity, while retroflex frication tend to generate lower CoG values as a result of less lingual constriction near the front oral cavity (Chuang & Fon, 2010). This allows us to effectively capture an overall measure of the degree of acoustic difference between frication at the two places of articulation. At the same time, CoG measurements are also sensitive enough to identify acoustic differences in the contrastiveness of alveolar and retroflex frication in speakers of different varieties of Mandarin. For instance, studies have shown that Taiwan Mandarin speakers tend to have smaller alveolar-retroflex contrast CoG differences as compared to Beijing Mandarin speakers, as well as higher retroflex CoG values consistent with the speech patterns of deretroflexion and phoneme contrast merger typically found in Taiwan Mandarin speakers (e.g., Chang 2012; Tso; 2017). Applying a CoG analysis on our participants' speech tokens will thereby allow us to clearly document an overview of the acoustic nature of the alveolar-retroflex contrast in Singapore Mandarin for the first time. This will allow us to determine if the two phoneme categories are in fact produced contrastively in contemporary Singapore Mandarin.

At the same time, while CoG values are a robust means of characterising Mandarin frication spectra, the acoustic data obtained from this measurement is condensed to a single measure. This analysis may therefore overlook more important details of the acoustic nature of the alveolar-retroflex contrast in Singapore Mandarin. To capture this fine-grained acoustic detail, we also examined the Long-Term Average Spectra (LTAS) of the two places of articulation. The LTAS is a fast Fourier transform-generated power spectrum which maps out the spectral density of frequencies across the full duration of a frication (Löfqvist & Mandersson, 1987). While the LTAS is commonly used to capture spectral densities across longer speech signals with vocal tract movement, it has currently been adopted as a robust method of modelling spectral densities across shorter duration noisy frication spectra in different languages (Heeren, 2015; Kendall & Fridland, 2021; Punišić et al., 2019; Stuart-Smith, 2020; van Dommelen, 2019). To model the normative spectral properties of our participants' alveolar and retroflex speech productions, we conducted a General Additive Mixed Models (GAMM) analysis on the LTAS of each of the word tokens obtained. This allowed us to identify the specific frequency windows of significant spectral difference in the alveolar-retroflex contrast of Singapore Mandarin. While LTAS data have been used to model frication spectra contrasts in Songyuan Mandarin (Li, 2008), this study will be the first to do so for the alveolar-retroflex frication contrast in Singapore Mandarin.

To promote open science practices, we have made all relevant data, task files, stimuli, and analysis files for this study available on the Open Science

Framework (https://osf.io/g4e2s/) and all speech tokens on DR-NTU (https://doi.org/10.21979/N9/ZTMPML).

4.2 Speech perception: Spring-Village word identification task

This word identification task was designed to test how accurate our participants were at distinguishing between minimal pairs of alveolar and retroflex initial words in both Singapore Mandarin and standard Beijing Mandarin.

4.2.1 Methods

4.2.1.1 Participants

The 50 participants who successfully completed our first distributional learning study in Chapter 2 were invited to, and agreed to participate in this study. The participants (39 female, 11 male) were aged between 18 – 25 years of age (*M* = 20.3, *SD* = 1.7). The majority of participants reported spending their entire lives in Singapore, with the exception of two participants who had spent portions of their lives in other countries – one in Malaysia during their preschool years, and one in Australia during their high school years. Informed consent was collected from all participants prior to commencement of the study, and this study was approved by the IRB of Nanyang Technological University institution (IRB-2019-01-034).

4.2.1.2 Equipment

The Spring-Village word identification task was conducted on a 10-inch LCD screen on a Microsoft Surface Go computer. All stimuli were presented in

OpenSesame (Mathôt et al., 2012). Audio stimuli were presented to the participants with a pair of Audio-Technica Over-ear Monitoring Headphones ATH-M40x.

4.2.1.3 Stimuli

Audio stimuli for this task consisted of two sets of 12 minimal pairs of word tokens – one set in Singapore Mandarin, and the other in standard Beijing Mandarin. These were the same words used in our ultrasound study (see Table 3.1 for the full list of words). The words were chosen following consultation with five fluent local Singapore Mandarin speakers. To ensure that this task would be well suited for participants across the lifespan and with different proficiencies in Mandarin, the selected words were understandable by school-aged children, and were concrete and imageable.

To create each audio word token for this task, audio recordings were made for each of the words on the wordlist by one Singapore Mandarin speaker, and on native standard Beijing Mandarin speaker. Each word was read three times by each speaker. All recordings were conducted in a sound attenuated recording booth, with a RODE NT1 Cardioid Condenser Microphone and a pop filter positioned approximately 30 cm from the speaker’s mouth. Recordings were recorded in stereo WAV format at a bit depth of 16 and sampling rate of 44.1 kHz using a Zoom H4n Pro Handy recorder. To extract each individual word token, tokens were spliced from each recording in Goldwave [v.6.51], and were saved in WAV format. The final token chosen for each word presented in the word identification task was selected on the basis of audio clarity. The audio tokens used for the task can be found in our online repository of study stimuli (https://osf.io/g4e2s/).

Visual stimuli for this task consisted of 12 picture cards depicting the six minimal pairs of words in the wordlist (see Table 3.1 for words). Each card featured an illustrated depiction of the meaning of the word, along with the Chinese character of the word, and its corresponding pronunciation in Hanyu Pinyin (see Figure 4.1 for example of a pair of cards). Hanyu Pinyin cues were included on each card to ensure that all participants would be able to read the words on the cards. The full set of illustrations and word picture cards can be found in our online repository (https://doi.org/10.21979/N9/ZTMPML).

4.2.2 Procedure

4.2.2.1 Word identification task

Participants were first introduced to the word identification task with written instructions in English. Participants were informed that they would be presented with an audio recording of a Mandarin word and a matching pair of picture cards on each test trial. Their task was to make a button press response to indicate which of the two picture cards contained the word that they heard.

Following the instructions, participants completed one block of word identification trials, with a total of 24 trials. On each trial participants were presented with one of the Mandarin words from the 24 word tokens in Singapore and Beijing Mandarin together with the corresponding minimal pair of picture cards. Written instructions in English were displayed on each trial to prompt participants to make a response. Figure 4.1 depicts an example of the visual stimuli presented on each trial.

Figure 4.1

Example of visual stimuli presented in one trial of the word identification task.

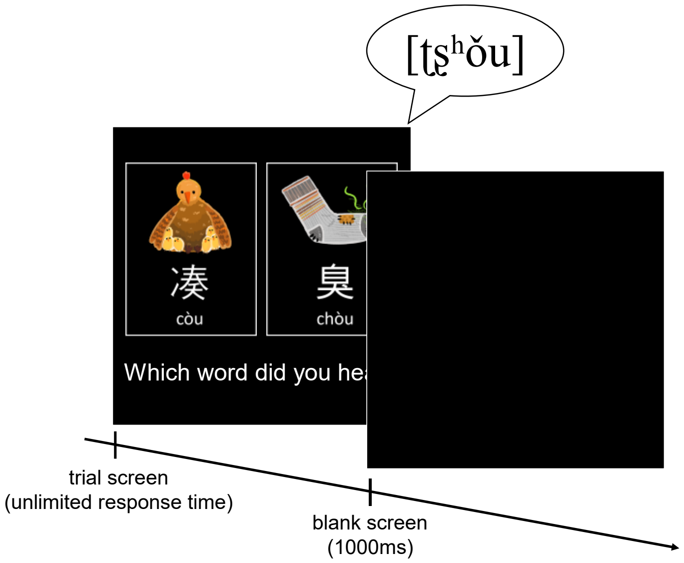
Which word did you hear?

Note. Figure depicts the words: [ʈsʰǒu] (left) and [ʈʂʰǒu] (right). Not drawn to scale.

Following a button press response, participants would hear a chime sound, and a blank screen would be presented for 1000ms. The participant would then be presented with the next trial (See Figure 4.2 for trial procedure). No feedback was given to the participants so as to not influence their response patterns. No trial timeout was implemented to ensure that participants had sufficient time to respond to all the words presented. The audio word tokens and their corresponding minimal pair of picture cards were presented in random order for each participant.

Figure 4.2

Example of Mandarin alveolar-retroflex Spring-Village word identification trial.



Note. Not drawn to scale.

4.2.2.2 Language ID

To obtain information on our participants' language repertoires, participants were asked to rate their ability in understanding the four official languages of Singapore (English, Mandarin, Malay, and Tamil), along with any other languages and dialects they understood on a Likert scale of 0, "I do not understand this language", to 7, "I have native-level understanding of this language".

4.2.3 Analysis plan

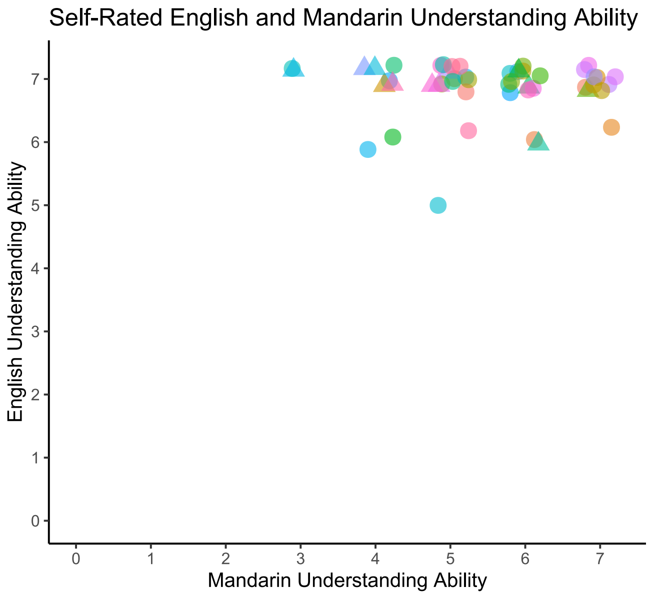
All analyses were conducted with R (R Core Team, 2020). To obtain word identification accuracy scores for each participant, we obtained the number of correctly identified alveolar words for all alveolar words presented, and the number of correctly identified retroflex words for all retroflex words presented. Accuracy scores for Singapore Mandarin and Beijing Mandarin were computed separately.

4.2.4 Results

4.2.4.1 Language ID

Figure 4.3

Participants' self-reported English and Mandarin understanding ability.



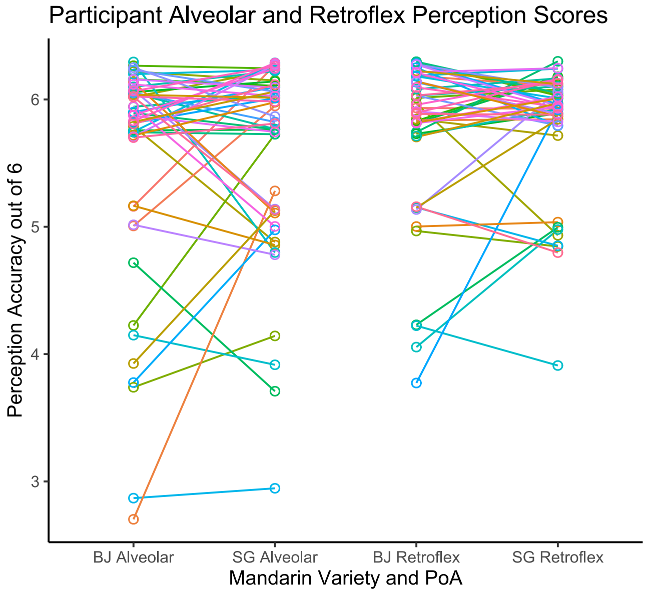
Note. $N = 50$. Circles: female, triangles: male. Responses jittered by .25 for visualisation.

Figure 4.3 shows the distribution of participants' self-rated English and Mandarin understanding abilities. Overall, our participants' self-reported English understanding scores were high (Range: 5-7, *Median* = 7, *SD* = 0.4), this indicating that they would not have had any issues in understanding the task instructions of the study. On the other hand, their Mandarin understanding scores showed more variability (Range: 3-7, *Median* = 5.5, *SD* = 1.1).

4.2.4.2 Word identification task.

Figure 4.4

Distribution of alveolar and retroflex perception scores for each variety of Mandarin.



Note. *N* = 50. Jittered by .3 for visualisation.

Participants' individual accuracy scores are presented in Figure 4.4. Visual inspection reveals that the majority of our participants performed at ceiling level accuracy at identifying alveolar and retroflex words in both Singapore Mandarin and Beijing Mandarin. Table 4.2 reports the summary statistics of the overall accuracy scores of all participants.

Table 4.1

Summary statistics for accuracy score on word identification task.

Variety of Mandarin	Place of Articulation	Median Accuracy Score (out of 6)	Accuracy Score Range
Singapore Mandarin	Alveolar	6	3 – 6
	Retroflex	6	4.2 – 6
Beijing Mandarin	Alveolar	6	3 – 6
	Retroflex	6	4 – 6

4.2.5 Discussion

Our word identification task was designed to examine differences in Singapore Mandarin speakers' perception of the Mandarin alveolar-retroflex contrast. We reasoned that we would be able to capture an objective measure of individual variance in perceptual sensitivity for this contrast using this task. However, the results showed that our participants had an overall high level of accuracy at distinguishing between the alveolar and retroflex places of articulation in both the local variety of Singapore Mandarin, as well as the standard variety of Beijing Mandarin.

The ceiling level accuracy we found for the local Singapore Mandarin word tokens is consistent with existing studies which have demonstrated that individuals typically show a good degree of perceptual sensitivity for the alveolar-retroflex contrast in a native variety of Mandarin. For instance, Tso (2017) found that Taiwan Mandarin speakers typically demonstrate high levels of accuracy at identifying alveolar and retroflex words spoken by other native Taiwan Mandarin speakers. On the other hand, the ceiling level accuracy we found for the Beijing Mandarin word tokens contrasts with the study by Shih and Kong (2011) which showed that some Taiwan Mandarin speakers demonstrate a degree of difficulty in distinguishing between alveolar retroflex words in a non-native variety of Beijing Mandarin. This suggests that speakers of contemporary Singapore Mandarin may have greater familiar with the phonology of standard Beijing Mandarin due to exposure to this variety of Mandarin in formal education (Chong & Tan, 2013).

Following this task, we set out to examine the acoustic nature of contemporary Singapore Mandarin speakers' production of the Mandarin alveolar-retroflex contrast. While early observational studies have suggested that there is an almost complete merger of the alveolar-retroflex phoneme contrast in Singapore Mandarin (e.g., Ng, 1985), the influence of standard Mandarin in formal education and Mainland Chinese media could have an impact on the degree of phoneme contrast present in the alveolar-retroflex speech productions of contemporary Singapore Mandarin speakers.

4.3 Speech production: Spring-Village speech production task

This speech production task was designed to collected data on our participants' speech productions of alveolar and retroflex initial words in

Singapore Mandarin. Speech elicitation was conducted with the use of the same picture cards used in the word identification task above, and speech recordings were obtained for fine-grained acoustic analyses to examine the acoustic nature of the alveolar-retroflex phoneme contrast in contemporary Singapore Mandarin.

4.3.1 Methods

4.3.1.1 Participants

The 50 participants who completed the word identification task were invited to then participate in the speech production task. One male participant was excluded from the acoustic analyses due to a failure to follow the task instructions. Additionally, as this study took place during a heightened period of contact restrictions during the COVID-19 pandemic, participants were not allowed to access the sound attenuated recording booth for this study. Therefore, although the audio recordings for this task were conducted in a quiet office room, occasional background noises were captured in the recordings. As the acoustic pattern of frication is easily masked by high ambient noise floors, some of the recordings were not suitable for fine-grained acoustic analysis. During each recording session, the researcher made a note of loud, unexpected background noises, together with the nature of the noise. To ensure that our audio data were of sufficient quality for the frication analyses we had planned, recordings were excluded from data analysis if they contained loud noises of thunderstorms, passing jet planes, air-conditioning noises, and voices of people talking outside the office room.

As CoG measurements are highly sensitive to deviations in background noise, 11 of the recordings were excluded from our CoG analyses due to background noises, leaving 39 suitable audio samples (30 female, 9 male) from participants aged between 18 – 24 years (*Median* = 20; *SD* = 1.5). On the other hand, we were able to retain more data for our GAMM analysis of the LTAS data as general additive mixed modelling is robust to frequency fluctuations in the sound signal and is better suited to account for unexpected noise artefacts in the audio data. Thus, eight recordings were excluded from the GAMM analysis due to background noises, leaving 42 suitable audio samples (32 female, 10 male) from participants aged between 18 – 25 years (*Median* = 19; *SD* = 1.6).

Informed consent was collected from all participants prior to commencement of the study, and this study was approved by the IRB of Nanyang Technological University institution (IRB-2019-01-034). We also collected consent from participants for their audio recordings to be released under the terms of the Growing Collection (see Appendix 4.1). All participants who took part in the production task consented to the Growing Collection, and chose a random identifier to be uploaded with their recordings.

4.3.1.2 Equipment

The Spring-Village speech production task was conducted with a 10-inch LCD screen on a Microsoft Surface Go computer. All stimuli were presented in OpenSesame (Mathôt et al., 2012). Speech recordings were made with a Zoom H4n Pro Handy Recorder. Participants spoke into a RODE

Lavalier lapel microphone attached to a RODE Lav-Headset mount. All recordings were made in stereo WAV format, with a sample rate of 44.1kHz and a bit depth of 16 bits.

4.3.1.3 Stimuli

Visual stimuli consisted of the same 12 Mandarin alveolar-retroflex word picture cards used in the word identification task. No audio stimuli were used for this speech production task.

4.3.2 Procedure

Participants were first introduced to the speech production task with written instructions in English. Participants were informed that a picture card would appear on the screen, and their task was to read each word aloud three times with a three second pause between each reading. To ensure that speech was produced as naturalistically as possible, participants were instructed to pronounce the words in a manner similar to that of daily speech. Emphasis was placed on the fact that the task was not a test of pronunciation accuracy, and that we were simply interested in documenting the unique speech patterns of Singapore Mandarin speakers.

Following the instructions, participants completed one block of audio recording consisting of 12 trials. For each trial, participants were presented with one of the 12 Mandarin alveolar-retroflex picture cards together with written instructions in English prompting them to read each word three times. Figure 4.5 shows an example of the visual stimuli presented in a trial. After reading each word three times, the participant made a keypress to move on to the next trial. The 12 picture cards were presented to each participant in random order. No feedback was given to the participants so as to not influence their speech production. No trial timeout was implemented to ensure that participants had sufficient time to read all the words presented.

Figure 4.5

Example of visual stimuli presented in one trial of speech production task.

Please read this word 3 times, then press the spacebar for the next word.

Note. Picture card depicts the word [ʈʂʰuə́n]. Not drawn to scale.

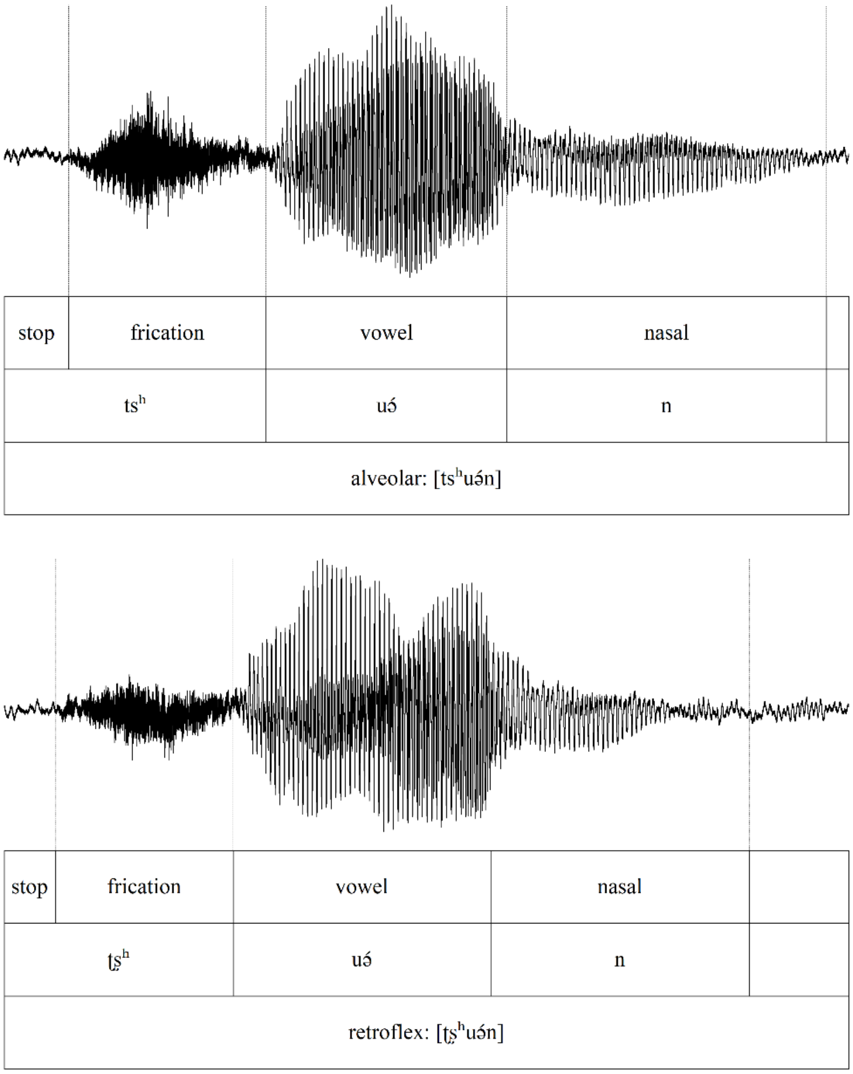
4.3.3 Analysis Plan

4.3.3.1 Audio Processing

Individual word tokens were obtained from each of the participants' audio files. Each word was manually selected, labelled, and spliced from the raw audio file in Goldwave [v.6.51], and saved as separate audio files in wav format. The complete set of raw audio files and individual word tokens are archived in the Spring Village Corpus under the Growing Collection of audio recordings (Goh et al., 2022) and can be accessed here: https://doi.org/10.21979/N9/ZTMPML. Following this, the frication portion of each segmented token was manually selected based on visual inspection of the waveform, labelled, and spliced in Praat 6.1.16 (Boersma & Weenink, 2020). Each individual frication token was saved as separate files in short txt and wav format. Figure 4.6 shows the identified frication portions of an alveolar word [tsʰuə́n] and a retroflex word [ʈʂʰuə́n] token produced by one participant (Participant 8). Figure 4.7 shows the spectral slices of the full spectra of the extracted frication of an alveolar word [tsʰuə́n] and a retroflex word [ʈʂʰuə́n] token produced by one participant (Participant 8).

Figure 4.6

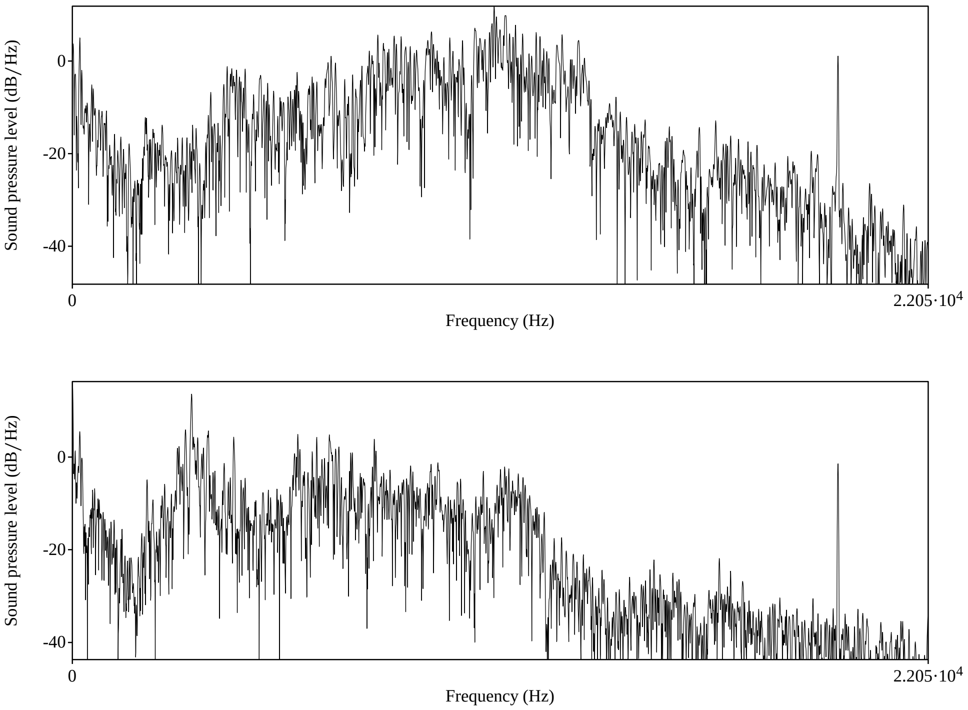
Waveform of alveolar [tsʰuə́n] (top) and retroflex word [ʈʂʰuə́n] (bottom).



Note. Both recorded from the same participant (Participant 8). Segmentation is conservative such that no voicing from the following vowel is included in the voiceless segment for analysis.

Figure 4.7

Spectral slices of frication spectrum of alveolar [tsʰuə́n] (top) and retroflex [ʈʂʰuén] (bottom).



Note. Both recorded from the same participant (Participant 8).

4.3.3.2 CoG Analysis

Following the methodology of Chang and Shih (2015) and Tso (2017), frication CoG values were obtained from the middle 30ms of the frication spectra of each token using the centre of gravity function in Praat 6.1.16 (Boersma & Weenink, 2020). Summary statistics were obtained for the alveolar and retroflex CoG values obtained in this speech production task.

To determine if there was a significant difference between the CoG values of alveolar and retroflex frication produced in Singapore Mandarin, a linear mixed effects model analysis was conducted on the CoG data in R (R Core Team, 2020) with the *lme* function from the *nlme* package (Pinheiro et al. 2022). We chose to include gender as a factor of interest in this analysis, as studies on Taiwan Mandarin have previously shown that retroflexion may be less common in male speakers due to extralinguistic factors (Rau & Li, 1994; Chuang & Fon, 2010). The CoG analysis therefore looked at the categorical fixed factors of place of articulation (alveolar vs retroflex) and gender (male vs female), as well as the interaction between place of articulation and gender, with random intercept for participant and random by-participant slope for trial number.

4.3.3.3 LTAS GAMM Analysis

To conduct the GAMM analysis, we exported the full frication spectra of each word token as LTAS matrices from Praat 6.1.16 16 (Boersma & Weenink, 2020). Using the *bam* function from the *mgcv* package (Wood, 2011) in R (R Core Team, 2020), we conducted a GAMM analysis on the LTAS matrices with the parametric fixed terms of place of articulation (alveolar vs retroflex) and gender (male vs female), along with a parametric interaction term between place of articulation and gender, smooth terms of place of articulation and gender, and random smooths for participant and trial number. Based on Sóskuthy (2017), a significant effect of a parametric term indicates a significant difference in the overall spectral density of the modelled data for each condition of the parametric term. Thus, we reported the effects of each parametric term assessed in the model. To assess for differences within the modelled spectral trajectories of each place of articulation, individual windows of spectral difference between the alveolar and retroflex places of articulation were identified using the *plot_diff* function from the *itsadug* package (van Rij et al., 2020) in R (R Core Team, 2020).

To follow up on any significant gender interaction effect, individual GAMMs were carried out on the alveolar and retroflex LTAS data separately with the parametric fixed term of gender (male vs female), along with smooth term of gender, and random smooths for participant and word token trial number. Individual windows of spectral difference between each gender for each place of articulation were then identified using the *plot_diff* function (van Rij et al., 2020) in R (R Core Team, 2020).

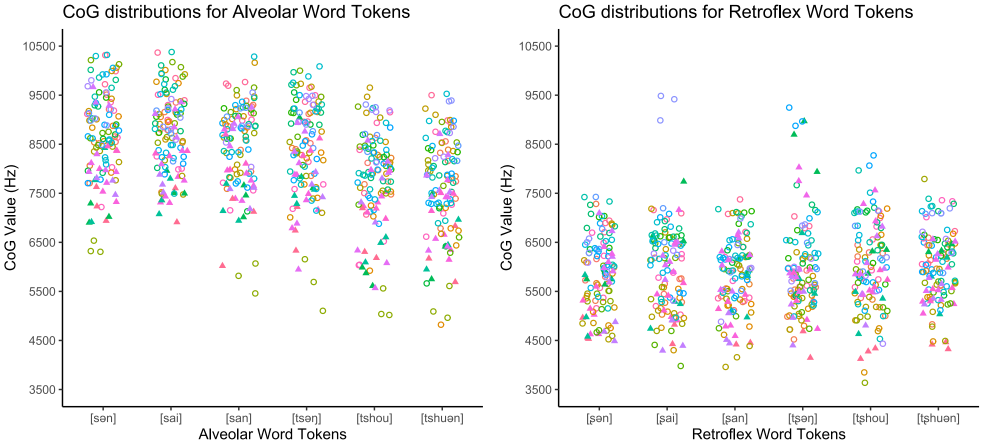
To ensure that the basis dimensions of each model would not result in oversmoothing of the LTAS data (Wieling, 2018), model criticism checks were conducted for each GAMM analysis using the *gam.check* function of the *mgcv* package (Wood, 2011). Based on the results of the criticism checks, the *k*-values of all models were set at 90. While this *k*-value is high, a comparison of the EDF values from the model criticism checks revealed that a *k*-value of 90 would be optimal for ensuring that an optimal amount of “wiggliness” of the LTAS data would be captured by the models.

4.3.4 Results

4.3.4.1 CoG Results

Figure 4.8

Distribution of each participant's alveolar and retroflex CoG values for each word token.



Note. *N* = 39. Circles: female, triangles: male.

Table 4.2

Summary statistics of alveolar and retroflex CoG values.

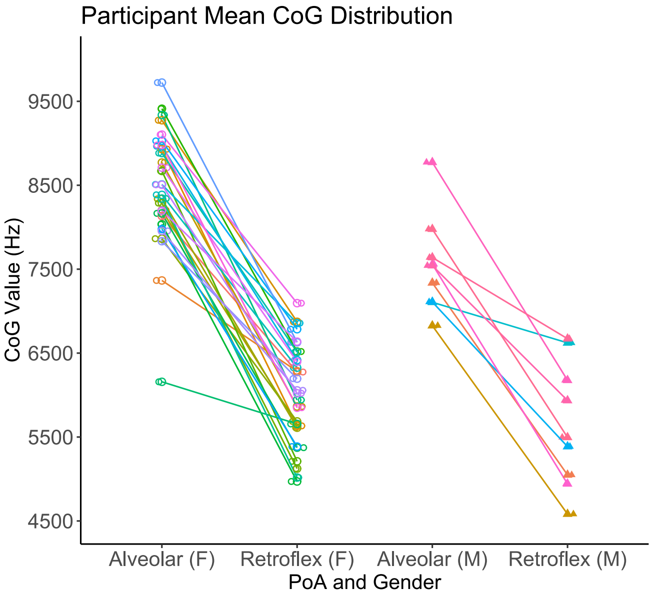
	Alveolar CoGs (Hz)		Retroflex CoGs (Hz)	
	Mean	Range	Mean	Range
Overall	8254.2 (1085.1)	2334.6 – 10678.4	5941.3 (852.7)	3637.2 – 9484.9
Female	8468.1 (1056.3)	2334.6 – 10678.4	6028.4 (822.1)	3637.2 – 9484.9

Male	7541.3 (852.5)	5568.0 – 9665.5	5651.3 (890.6)	4125.2 – 8970.4

Note. Standard deviation in parentheses.

Figure 4.9

Distribution of the mean CoG values of each participant by gender and PoA.



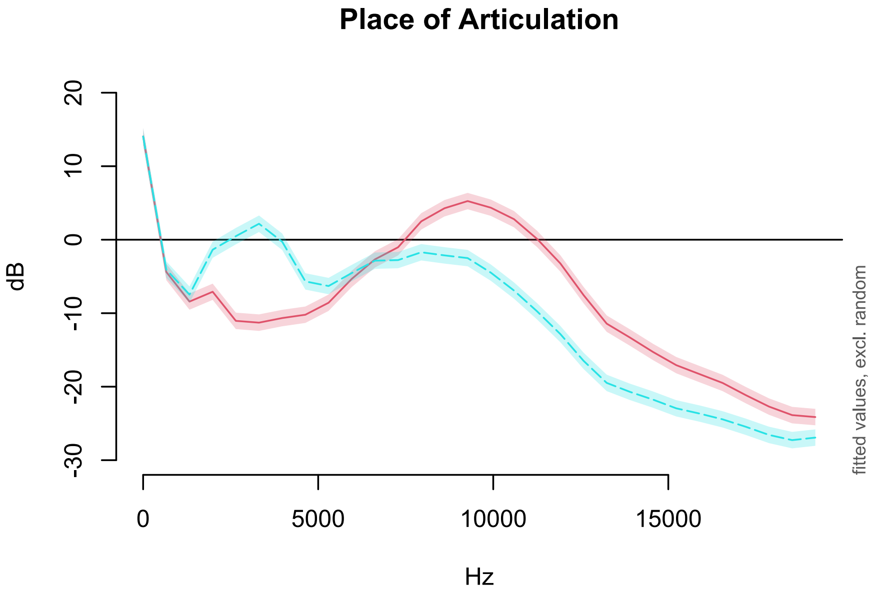
Note. $N = 39$. Circles: female; triangles: male.

Figure 4.8 shows the CoG values obtained from each individual alveolar and retroflex word token of each participant. Overall summary statistics for the CoG values are presented in Table 4.3. The linear mixed effects model analysis conducted on the CoG values revealed that there was a significant main effect of place of articulation, $t(1,1363) = -51.7$, $p < .0001$, $\eta p^2 = .70$ indicating that there was a significant difference between our participants' overall CoG values for their alveolar and retroflex frication, with higher CoG values for alveolar words, and lower CoG values for retroflex words (see Table 4 for summary statistics). The analysis also revealed a significant main effect of gender, *t*(1,37) = -4.38, *p* =.0001, ηp^2 = .21, indicating that there was an overall difference in the CoG values of our male and female participants, with female participants having higher overall CoG values as compared to male participants (see Table 4 for summary statistics). Finally, the CoG analysis also revealed a significant interaction between place of articulation and gender, *t*(1,1363) = 5.94, *p*<.0001, ηp^2 = .02. Figure 4.9 presents the mean CoG values of each participant by place of articulation and gender. A visual inspection of Figure 4.9 shows that male participants have lower CoG values than female participants, particularly in their alveolar frication, resulting in less contrast between the alveolar and retroflex places of articulation for male as compared to female participants.

4.3.4.2 LTAS GAMM Results

Figure 4.10

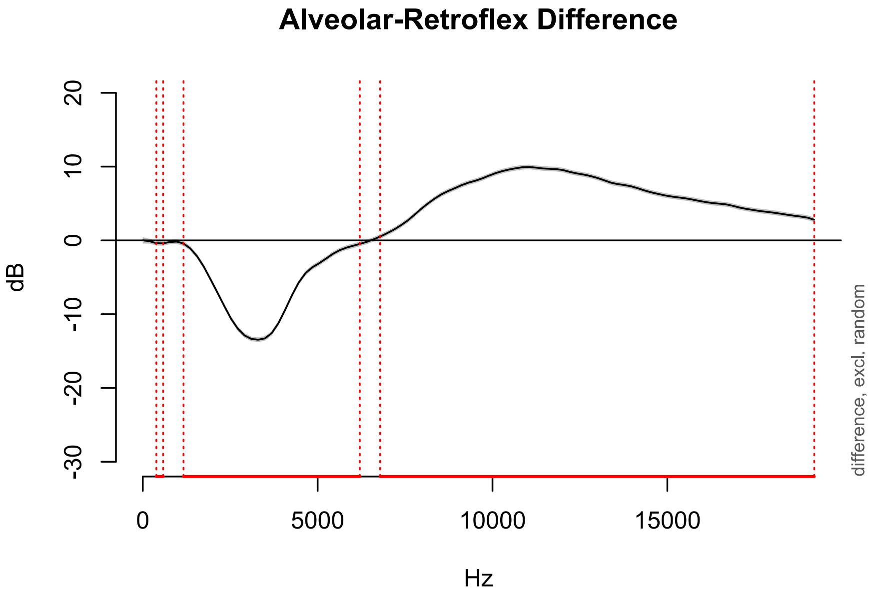
Fitted models of alveolar and retroflex frication LTAS data.



Note. Alveolar: red line; retroflex: turquoise dotted line.

Figure 4.11

Model of spectral difference between alveolar and retroflex frication LTAS data.



Note. Windows of significant difference bounded in red dotted lines.

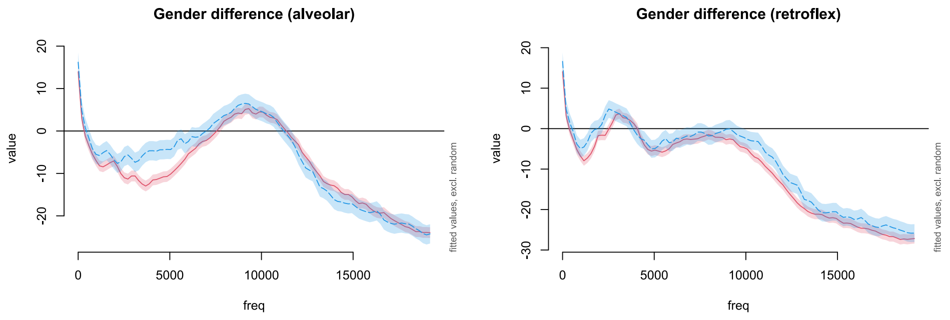
The overall fitted LTAS models of our participants' alveolar and retroflex frication are presented in Figure 4.10, plotted with the *plot_smooth* function from the *itsadug* package (van Rij et al., 2020). The individual frequency windows of significant difference between the alveolar and retroflex fitted models are presented in Figure 4.11, plotted with the *plot_diff* function from the *itsadug* package (van Rij et al., 2020).

The main GAMM analysis revealed that there was a significant main effect of the parametric term of place of articulation, *t*(1,1279) = -88.1, *p*

<.00001, indicating that there was a significant difference in the overall spectral densities of the alveolar and retroflex frication of our Singapore Mandarin speaking participants. The results also revealed a significant interaction between the parametric terms of place of articulation and gender, $t(1,1279) = 13.4$, $p < .00001$. No significant main effect of the parametric term of gender was found, $t(1,1279) = 1.12$, $p = .26$. Three frequency windows of significant difference between the modelled alveolar and retroflex spectra were found with the *plot_diff* function: 387.9 – 581.8 Hz, 1163.6 – 6206.1 Hz, and 6787.9 – 19200.0 Hz (as shown in Figure 4.11). This indicates that there were significant differences in not only the overall spectral density, but also in the shape of the modelled trajectories of the alveolar and retroflex frication spectra as produced by our Singapore Mandarin speaking participants.

Figure 4.12

Fitted models of alveolar and retroflex frication LTAS, coloured by gender.



Note. Alveolar models: left; retroflex models: right. Female: red solid lines; male: blue dotted lines.

As the main GAMM analysis revealed a significant interaction between the parametric terms of place of articulation and gender, two follow-up GAMM analyses were then conducted on the alveolar and retroflex LTAS data separately. These analyses revealed that there was no overall significant effect of the parametric term of gender on the alveolar, *t*(1,1219) = 0.98, p = 0.33, or retroflex, *t*(1,1219) = 1.92, *p* = .05 LTAS data. This indicates that there was no significant difference in the overall spectral densities of our male and female participants for either the alveolar or retroflex places of articulation. The fitted models for gender differences in the alveolar and retroflex LTAS data are shown in Figure 4.12, plotted with th/ *plot_smooth* function from the *itsadug* package (van Rij et al., 2020). However, three individual frequency windows of significant difference between each gender were found with the *plot_diff* function for both the alveolar and the retroflex spectra. Alveolar: 1336 –

1559Hz, 2450 – 6236Hz, and 14032 – 14255Hz, retroflex: 1114 – 1782Hz, 2227 – 2673Hz, and 10245 – 12918Hz. Moreover, visual inspection of Figure 4.12 reveals that while the spectral trajectories of both male and female participants were very similar for the retroflex place of articulation, the alveolar spectral trajectories of each gender diverged visibly at around the 2000 – 6000Hz region. This indicates that while there was little impact of gender on the overall spectral densities of each place of articulation, individual windows of spectral trajectory difference for each gender (particularly in the case of the alveolar place of articulation) could have contributed to the significant parametric interaction effect found between place of articulation and gender as found in the main GAMM analysis.

4.3.5 Discussion

While early observational studies in the 1980's have suggested that there is a high prevalence of deretroflexion in Singapore Mandarin (e.g., Lock, 1989; Ng, 1985), the results of our speech production task revealed evidence of a a significant acoustic contrast in articulations of alveolar and retroflex fricative phonemes in contemporary Singapore Mandarin. This acoustic contrast is evident in the results of both our CoG analysis as well as the LTAS GAMM analysis we conducted on a large sample size of adult Singapore Mandarin speakers. In particular, the LTAS GAMM results showed that there is not only a significant difference in the overall spectral densities of the two places of articulation, but that there is also evidence of significant differences in the spectral trajectory of each place of articulation with three specific frequency windows of difference being identified between the modelled alveolar and retroflex spectra.

The results we obtained have given us a clearer understanding of the acoustic nature of the alveolar-retroflex phoneme contrast in contemporary Singapore Mandarin, and we are now in a better position to understand how the local variety of Mandarin compares to other previously documented varieties of Mandarin. For instance, a documentation by Chang (2012) on the alveolar-retroflex contrast in Taiwan Mandarin and standard Beijing Mandarin speakers reported the CoG values of Taiwan Mandarin (Alveolar: *Mean* = 8757.6; Retroflex: *Mean* = 6224.1) and Beijing Mandarin (Alveolar: *Mean* = 9115.6; Retroflex: *Mean* = 5160.7) speakers. While the sensitive nature of CoG analysis means that we cannot directly compare our results to this data, our CoG results nevertheless suggests that the articulations of our participants may be closer in nature to that of Taiwan Mandarin speakers as compared to standard Beijing Mandarin speakers (see Table 4 for summary statistics of our CoG values). This suggests that there may still be a degree of alveolar-retroflex phoneme contrast merger in Singapore Mandarin, although not to the extent that was observed in the 1980's (e.g., Lock, 1989; Ng, 1985).

At the same time, we have also identified a possible effect of gender on alveolar-retroflex articulations in Singapore Mandarin. Our CoG analysis revealed that male participants produced lower overall CoG values (particularly in their alveolar frication), and less contrast between their alveolar and retroflex

CoG values. The LTAS GAMM analysis also revealed that male participants produced a higher concentration of lower frequencies in their alveolar frication, particularly in the 2000 – 6000Hz region (see Figure 4.12). It is possible that the cause of this difference could be anatomical in nature due to structural differences in male and female vocal tracts – for instance, the longer front oral cavities of male vocal tracts have been found to contribute to lower spectral peaks in frication (e.g., Jongman et al., 2000). On the other hand, it is also possible that our male participants could have been intentionally using different articulations to approximate more “formal” acoustic targets. This could be akin to the phenomenon of hypercorrection which has been documented in Taiwan Mandarin speakers. Chung (2006) found that Taiwan Mandarin speakers would engage in hypercorrection when they were aware that retroflexion should be present in Mandarin speech, yet remain uncertain of which words should be retroflexed, leading to overcompensation in retroflexion to account for possible lapses in speech. Indeed, a study conducted on Singapore Mandarin speakers in the 1980s found that male speakers were more likely to engage in hypercorrection as compared to female speakers in contexts designed to elicit more ‘formal’ speech production (Ng, 1985). At this point in time, however, we cannot make any generalisations to the larger population of Singapore male speakers as our sample size of male participants in this study is very small (CoG: N = 9; LTAS: N = 10). Future studies can focus on collecting more data from male speakers to obtain a more representative sample size.

4.4 Conclusion

This study has presented us with novel information on the nature of the alveolar-retroflex phoneme contrast in contemporary Singapore Mandarin. To our knowledge, this is the first time in which this contrast in the local variety of Mandarin has been documented in a large sample size with the use of fine-grained acoustic analyses and a test of word identification accuracy. The results of our study demonstrate that Singapore Mandarin speakers do indeed distinguish between the two places of articulation, both in production as well as in perception. While the place distinction observed in our participants' speech production is not as contrastive as in Beijing Mandarin, our participants nevertheless showed high levels of accuracy at perceptually distinguishing words in minimal pairs of both Singapore and standard Beijing Mandarin. These results may reflect shifting attitudes towards standard Beijing Mandarin as a prestige language used in formal education in Singapore (Chong & Tan, 2013).

Overall, the results of our study contribute not only to local understanding of Singapore Mandarin, but also add to a growing body of work on the phonology of outer-circle varieties of languages. In the local context of education in Singapore, our acoustic analyses can help to inform the way in which Mandarin is taught and assessed. The speech patterns of our participants demonstrate that there can be considerable variation in production of the alveolar-retroflex contrast in Singapore Mandarin, while not precluding high levels of comprehension for words in both the local and standard Beijing varieties of Mandarin. Understanding the local articulations of Mandarin speech can enable educators to account for the speech norms of local students in lessons and assessments, while also helping them to understand how phonology varies in other varieties of Mandarin.

The documentation of this outer-circle language of Singapore Mandarin is also of importance in contributing to variationist sociolinguistics perspectives on the legitimisation of variations in language use in different contexts. Moreover, our study also provides new insight into the effects of contact-induced change in an outer-circle language over the course of a few decades – while the early use of Singapore Mandarin was highly influenced by non-retroflex Chinese languages, the effect of more recent contact with standard Beijing Mandarin can already be seen in the retroflexion found in our participants. The availability of acoustic norms and corpora of outer-circle languages is also crucial for well-designed studies in the fields of language cognition and psycholinguistics. Such documentation will ensure that researchers will have access to language stimuli that is aligned with the acoustic norms of local varieties rather than having to resort to the use of "standard" varieties that may differ significantly from local phonological norms. At the same time, acoustic documentation of outer-circle varieties of languages is also crucial for ensuring the validity of medical tests of hearing. Speech audiometry tests are an important component of audiologic tests, where they are used for identifying and diagnosing hearing impairments specific to different phonological aspects of speech. Studies have shown that it is necessary for speech audiometry material to be phonologically aligned with the native variety of the listener in order to ensure that results are not confounded by stimuli that is inherently unfamiliar to the listener (Soh & Loo, 2023). This can be particularly difficult to accomplish in the case of outer-circle varieties of languages that lack sufficient acoustic documentation such as in Singapore Mandarin. As a result of this, Soh and Loo (2020) have found that there is no existing Mandarin speech audiometry material suitable for accurately assessing hearing in Singapore Mandarin speakers due to the confounding effects of word, accent, and phonological unfamiliarity in audiometry batteries of Taiwan and standard Mandarin. Our acoustic documentation in this study can function as a step in informing the creation of suitable Mandarin audiometry material that is aligned to the acoustic norms of contemporary Singapore Mandarin.

From this study, we can see that the phonology of the alveolar-retroflex contrast in Singapore Mandarin differs from other well-documented varieties of Mandarin with existing corpora (e.g., The Chinese Lexicon Project, Sze et al., 2014; The Chinese Lexicon Project, Tse et al., 2017). In order to accurately assess a language-specific transfer effect in our target population of Singapore English-Mandarin bilinguals, we thus synthesised our own training stimuli based on the acoustic norms of the local variety of Singapore Mandarin obtained in this study. As the perception task used in this study showed that word identification alone was not sensitive enough to capture individual variance in perception of the Mandarin alveolar-retroflex contrast, we designed a new study using this new set of stimuli to capture a more fine-grained measure of individual differences in perception for a Singapore Mandarin alveolar-retroflex contrast.

Chapter 5: Fine-grained analysis of perception of a Singapore Mandarin alveolar-retroflex sibilant contrast (Spring-Village CROWN Game)

5.1 Introduction

5.1.1 Synthesis of novel Singapore Mandarin alveolar-retroflex stimuli

Our earlier documentation of the Singapore Mandarin alveolar-retroflex contrast show that the phonology and articulation of Singapore Mandarin speakers appears to differ from other varieties of Mandarin. To test for the presence of language-specific transfer effects in Singapore English-Mandarin bilingual adults, we would need speech sound stimuli that share structural overlaps with the unique local variety of Singapore Mandarin. We reasoned that stimuli in existing Mandarin corpora such as The Chinese Lexicon Project (Sze et al., 2014) and the two-character compound word Chinese Lexicon Project (Tse et al., 2017) would not be best suited for our purposes as they are based on the speech of standards of Mainland Mandarin and Hong Kong Mandarin speakers whose articulatory norms likely differ from Singapore Mandarin speakers. As such, we synthesised our own novel set of Mandarin alveolar-retroflex stimuli based on the acoustic norms of contemporary Singapore Mandarin speech as assessed in Chapter 4.

The stimuli that we created consists of a synthesised 8-step Singapore Mandarin [tsʰuə́n]-[ʈʂʰuə́n] (村 ‘village’ - 春 ‘spring’) alveolar-retroflex continuum contrast aligned to the acoustic measurements of the local Singapore Mandarin speakers in our earlier documentation. Our decision synthesise this stimulus set with the minimal word pair [tsʰuə́n]-[ʈʂʰuə́n] was based on the following criteria. First, we needed a minimal pair of words that are imageable and easily recognisable by Singapore Mandarin speakers. Second, we specifically needed a word pair that is perceptually difficult to discriminate. Our study in Chapter 3 showed that Singapore Mandarin speakers tend to have very high levels of accuracy at discriminating minimal pairs of alveolar and retroflex words. Thus, perceptual complexity would be necessary to capture a fine-grained measure of individual differences in speech sound perception. Third, we needed to ensure that the word pair we chose would contain a Singapore Mandarin alveolar-retroflex contrast that unique to Mandarin, and is not similar to any speech sounds present in Singapore English.

The minimal pair [tsʰuə́n]-[ʈʂʰuə́n] fit all three of these criteria. The words are easily recognisable and imageable with the alveolar [tsʰuə́n] meaning “village”, and the retroflex [ʈʂʰuə́n] meaning “spring” (the season). Next, the low rounded vowel present in both words lends perceptual complexity to this contrast. Existing studies show that there is a larger degree of phonological ambiguity for Mandarin alveolar and retroflex frication that are followed by low rounded vowels (Chang, 2012; Jeng, 2006). Our study in Chapter 4 confirms this, with a visual inspection of Figure 4.8 showing that a larger degree of alveolar-retroflex CoG overlap for the minimal pair [tsʰuə́n]-[ʈʂʰuə́n] as compared to the words. Finally, the alveolar affricate /tsʰ/ is not used contrastively with a retroflex or post-alveolar phoneme in a word-initial place in Singapore English. This would allow us to avoid any potential spillover effects from the phonology of Singapore English, as it is possible that there may be overlaps between the Singapore English alveolar-post-alveolar /s/-/ʃ/ phoneme contrast, and the Singapore Mandarin /s/-/ʂ/ phoneme contrast. However, the alveolar-retroflex affricate contrast /tsʰ/-/ʈʂʰ/ is found only in Mandarin but not English. Thus, we reasoned that the word pair [tsʰuə́n]-[ʈʂʰuə́n] would be ideal for our stimulus set.

To create our 8-step Singapore Mandarin [tsʰuə́n]-[ʈʂʰuə́n] alveolar-retroflex contrast continuum, we first used a GAMM analysis to model the overall averaged spectral trajectories of all the [tsʰuə́n] and [ʈʂʰuə́n] frication LTAS matrix files obtained from the 42 participants in our last study with a GAMM analysis. A Euclidean distance analysis of similarity was then conducted with all the individual [tsʰuə́n] and [ʈʂʰuə́n] LTAS matrix files to identify the word tokens that had the smallest Euclidean distance from the modelled overall average spectral trajectories. Following this, a trained native Singapore Mandarin speaker recorded several high audio quality tokens of the words [tsʰuə́n] and [ʈʂʰuə́n] in a sound attenuated recording booth while listening to the [tsʰuə́n] and [ʈʂʰuə́n] word tokens identified with the Euclidean distance analysis. All tokens were recorded with a Zoom H4n Pro Handy Recorder, and the speaker spoke into a microphone with a pop filter. All recordings were made in stereo WAV format, with a sample rate of 44.1kHz and a bit depth of 16 bits.

The most suitable high audio quality [tsʰuə́n] and [ʈʂʰuə́n] word tokens were then selected with a second Euclidean distance analysis of similarity to the modelled overall average [tsʰuə́n] and [ʈʂʰuə́n] frication spectra of the earlier 42 participants. We then ensured that the chosen high audio quality word tokens also sounded as similar to each other as possible in terms of pitch, vowel transitions, vowels, and nasals. We modified the durations of their initial frication onsets (i.e., before peak frequencies) and nasal offsets (i.e., after peak frequencies) with Goldwave [v.6.51] to ensure that both tokens were similar in length, and then equalised the intensity and pitch tracks of both tokens in Praat (Boersma & Weenik, 2020). This ensured that perceptual ambiguity between these two tokens would be isolated to the frication spectra as much as possible.

Once we had obtained our two final high audio quality [tsʰuə́n] and [ʈʂʰuə́n] word tokens, we synthesised our 8-step continuum using these two tokens as the continuum endpoints. This was conducted with the TANDEM-STRAIGHT speech analysis and synthesis software (Kawahara & Morise, 2011) in MATLAB (R2020a), following the methodology of the sibilant continuum creation tutorial designed by McAuliffe (2017). This resulted in a synthesised [tsʰuə́n]-[ʈʂʰuə́n] Singapore Mandarin alveolar-retroflex contrast continuum of eight 580ms word tokens that were equally spaced from each other in acoustic distance. Our synthesised continuum stimulus set can be found here (https://osf.io/rnq4w). As this stimulus set was synthesised based on the norms of a large dataset of local Mandarin speakers, we reasoned that it would be well suited for examining language-specific transfer effects in Singapore

Mandarin speakers with a distributional learning task. However, before conducting this investigation, we first needed to ensure that the new stimulus set was indeed suitable for capturing fine-grained levels of individual differences in perceptual sensitivity among the local population of Singapore Mandarin speakers. To do so, we designed a new study paradigm to assess perceptual sensitivity for this new [tsʰuə́n]-[ʈʂʰuə́n] Singapore Mandarin alveolar-retroflex contrast continuum.

5.1.2 Spring-Village CROWN Game

We designed a new paradigm to obtain a more fine-grained measure of phoneme contrast perception – the Spring-Village CROWN Game. The Spring-Village CROWN Game is based on a 2-alternative forced choice (2AFC) task created by Ke et al. (2021) known as the CROWN Game. In this paradigm, participants are presented with sound tokens drawn from a continuum of speech sounds varying in acoustic ambiguity from each other. Participants are then instructed to make categorical identification decisions for each of the sounds they hear. The response patterns of each participant can then be used to measure perceptual sensitivity to different degrees of acoustic ambiguity in the sound tokens. Similar 2AFC paradigms have successfully been used to examine individual differences in phoneme contrast perception, such as in Chang (2012) who tested Taiwan and Beijing Mandarin speakers on their perceptual response patterns for a continuum of Mandarin alveolar-retroflex speech tokens (Chang, 2012). The original CROWN Game has also been successfully used to capture fine-grained individual differences in the perception of voice onset time (VOT) phoneme contrasts in English and Mandarin with multilingual participants (Ke et al., 2021; Pan et al., 2021). Thus, we were confident that this paradigm would be well suited for obtaining a more detailed measure of perceptual differences in the resolution of Mandarin alveolar-retroflex ambiguity in our target population of Singapore English-Mandarin bilinguals.

5.1.3 Link between fine-grained measure of sensitivity and Mandarin understanding ability

Besides using this new task to determine if our new Singapore Mandarin stimulus set could capture individual differences in perceptual sensitivity for alveolar-retroflex contrast ambiguity, we also planned to explore the question of whether this fine-grained measure of phoneme contrast perception might be linked to our participants' overall Mandarin understanding abilities. While other studies have yet to examine a link between Mandarin alveolar-retroflex phoneme contrast sensitivity and general Mandarin comprehension ability, we have some reason to believe that such a link may exist. Existing studies appear to show that there is a link between the manner in which the Mandarin alveolar-retroflex contrast is perceived, and the variety of Mandarin that individuals are familiar with. For instance, Chang (2012) compared how Taiwan and Beijing Mandarin speakers resolved perceptual ambiguity for sound tokens on an 8-step alveolar-retroflex contrast continuum synthesised from standard Beijing Mandarin. The results of the study showed that the variety of Mandarin that individuals spoke was linked to different patterns of perceptual ambiguity resolution. In particular, speakers of the deretroflexed variety of Taiwan

Mandarin displayed a higher level of tolerance for phonological ambiguity in retroflexion as compared to Beijing Mandarin speakers who had a stricter cut-off for the degree of phonological ambiguity in acceptable examples of retroflexion (Chang, 2012). It is possible that this difference in perception may extend to Mandarin comprehension, as Taiwan Mandarin speakers have been documented to display a degree of perceptual confusion when tasked with identifying alveolar and retroflex words in Beijing Mandarin speech (Shih & Kong, 2011). However, it is important to note that the test stimuli used in both of these studies consisted of word tokens aligned to the phonological norms of the standard Beijing variety Mandarin. Thus, the results found in these studies may not be generalisable to how Mandarin alveolar-retroflex phoneme contrast sensitivity is linked to language perception and comprehension in a local, familiar variety of Mandarin. In fact, studies have found that Taiwan Mandarin speakers typically show high levels of overall Mandarin comprehension ability despite some level of perceptual confusion for the alveolar-retroflex contrast in both the Taiwan and the standard Beijing varieties of Mandarin (e.g., Chung, 2006). Therefore, while the results of this study might reveal a link between overall Mandarin understanding abilities and perceptual sensitivity for the Mandarin alveolar-retroflex contrast, it is also possible that the results will show that high levels of daily Mandarin proficiency may be possible without the presence of correspondingly high levels of fine-grained perceptual sensitivity for the Mandarin alveolar-retroflex contrast.

To promote open science practices, we have made all relevant data, task files, stimuli, and analysis files for this study available on the Open Science Framework (https://osf.io/rnq4w/).

5.2 Methods

5.2.1 Participants

We had no prior estimate of effect size for our examination of a link between Mandarin understanding abilities and perceptual sensitivity for the Mandarin alveolar-retroflex contrast. As such, we preregistered a sample size range instead, with a Bayesian inference threshold as our data collection stop rule. Our minimum sample size was based on an a priori power analyses in G*Power 3.1.9.7 to observe a large effect size ($\rho = 0.5$) in a test of correlation at an alpha level of 0.05 with a power of $1-\beta = 0.95$. This analysis revealed that a minimum sample size of $N = 34$ participants would be required if a large effect size were present in the data. Our maximum sample size was based on an a priori power analyses in G*Power 3.1.9.7 to observe a minimum effect size of interest ($\rho = 0.3$) in a test of correlation at an alpha level of 0.05 with a power of $1-\beta = 0.95$. This analysis revealed that a sample size of $N = 111$ would be required if a minimum effect size were present in the data. Thus, we preregistered a sample range of $N = 34 – 111$. Once the minimum sample size of $N = 34$ was reached, a Bayesian significance test was then carried out on the data with each consecutive participant to determine if the results of a correlation analysis between self-rated Mandarin Understanding scores and Decision Gradient slopes (as described in the analysis section) were substantially supportive of either H1 (i.e., that there is a correlation between the two factors; BF>3), or H0 (i.e., that there is no correlation between the two factors; BF<0.33) (Dienes, 2014). The prior for the Bayesian significance test was set at a correlation of ρ = 0.25, based on a study conducted by Pan et al. (2021), and we would stop data collection once a substantial Bayes factor was obtained.

Based on our preregistered sample size guidelines, 69 participants (41 female, 28 male) were recruited from the student population of Nanyang Technological University in exchange for course credits or payment of $5 following completion of the study. Four female and three male participants were excluded from data analysis due to a failure to follow task instructions, or for not being from the target demographic of Singapore Mandarin speakers. The remaining sample of 62 participants were aged between 18 – 29 years of age (*Median* = 22.0, *SD* = 2.1), all of whom reported spending their entire lives in Singapore. Informed consent was collected from all participants prior to commencement of the study, and this study was approved by the IRB of Nanyang Technological University institution (IRB-2019-01-034).

5.2.2 Equipment

The Spring-Village CROWN Game was conducted on a 10-inch LCD screen on a Microsoft Surface Go computer. All stimuli were presented in OpenSesame (Mathôt et al., 2012). Audio stimuli were presented to the participants with a pair of Audio-Technica Over-ear Monitoring Headphones ATH-M40x.

5.2.3 Stimuli

Visual stimuli consisted of cartoon depictions of a monkey, and of the two target words in Mandarin Chinese: a village for [tsʰuə́n], and a spring festival (also known as Lunar New Year) for [ʈʂʰuə́n]. Figure 5.1 contains an example of the visual stimuli presented to participants.

Audio stimuli consisted of eight 580ms acoustic tokens synthesised along an 8-step Singapore Mandarin alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] (村 ‘village’ - 春 ‘spring’) continuum. Each token on the continuum was spaced equally from its adjacent tokens terms of acoustic distance. The alveolar end of the continuum (i.e., the most alveolar [tsʰuə́n]-like sound) was represented by Sound 1, and the retroflex end of the continuum (i.e., the most retroflex [ʈʂʰuə́n]-like sound) was represented by Sound 8.

5.3 Procedure

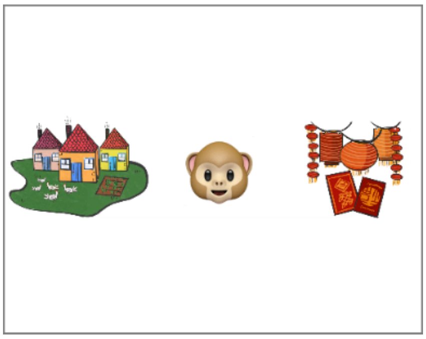
Participants were introduced to the task with written instructions in English. The instructions followed a storyline in which participants were told they needed to help a little monkey collect flowers for a flower crown by directing the little monkey travel to one of two locations (i.e., to the village or to the spring festival) based on the word that they heard on each trial.

5.3.1 Practice phase

To ensure that participants were familiar with the word sound to picture pairing, participants first completed two practice blocks. On each practice trial, participants were presented with the little monkey in the middle of the screen, and the pictures of the two locations (i.e., the village and the spring festival) on either side of the screen, while either Sound 1 or Sound 8 was played (see Figure 5.1 for example of visual stimuli). The participant would then make a keypress response to direct the monkey to either side of the screen. Following a keypress response, written feedback in English would be provided at the top of the screen (see Figure 5.2 for example of feedback provided following each practice response). Each sound token was played two times in random order over the course of each block, and the location of the village and spring festival pictures were counterbalanced across practice blocks.

Figure 5.1

Visual stimuli presented to participant in CROWN game.



Note. Not drawn to scale.

Figure 5.2

Examples of visual feedback given during training phase.

Note. Left: correct Sound 1 response. Middle: Correct Sound 8 response. Right: Incorrect response. Not drawn to scale.

5.3.2 Test phase

Following completion of the practice phase, participants completed a total of 10 test blocks. The test blocks were similar to the practice phase, with the exception that all eight sound tokens from the 8-step Singapore Mandarin alveolar-retroflex continuum were played in random order across each block, and no written feedback was given following a keypress response so as to not influence the participants’ response patterns.

5.3.3 Language ID

To collect data on participants’ language understanding abilities, they were asked to rate how well they understood the four main languages in Singapore (English, Mandarin, Malay, and Tamil) on a scale of 0 to 100 using an onscreen sliding scale. Participants were also given an option to add in more languages or dialects to rate.

5.4 Analysis Plan

5.4.1 Spring-Village CROWN GAME

We obtained two measures of Mandarin alveolar-retroflex contrast perception from the response patterns of our participants in the Spring-Village CROWN game. First, we determined the average boundary threshold on the 8-step Mandarin alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] continuum at which the majority of the participants’ sound token identification decisions switched from the alveolar-initial word [tsʰuə́n] to the retroflex-initial word [ʈʂʰuə́n]. Next, we obtained the slope value of the decision gradients of each participants’ response patterns for each sound token on the continuum. This decision gradient slope values represent the manner in which participants resolve perceptual ambiguity for speech tokens falling between the prototypical alveolar Sound 1 [tsʰuə́n] and retroflex Sound 8 [ʈʂʰuə́n]. Steep slopes represent a highly systematic response pattern consistent with a high degree of contrast sensitivity for perceptually ambiguous sound tokens the continuum, and shallow slopes represent a less systematic response pattern consistent with a lower degree of contrast sensitivity for perceptually ambiguous sound tokens on the continuum.

To obtain the boundary threshold and decision gradient slope values, we obtained the number of “spring festival” (Sound 8 retroflex [ʈʂʰuə́n]) responses each participant made for each of the eight sound tokens on the alveolar-retroflex contrast continuum across the test phase. Psychometric functions were then fit for each participant with *quickpsy* function in R (Linares & López-Moliner, 2016; R Core Team, 2020) based on the proportion of “spring festival” (Sound 8 retroflex [ʈʂʰuə́n]) responses made out of all the responses for each of the eight tokens on the alveolar-retroflex contrast continuum. Individual slope and boundary threshold values were then extracted for each participant from their fitted psychometric functions. Following this, we then obtained the median boundary threshold of all our participants. This median boundary threshold represents the median Mandarin alveolar-retroflex boundary threshold for our Singapore Mandarin speakers.

5.4.2 Link between Mandarin understanding and decision gradient slopes

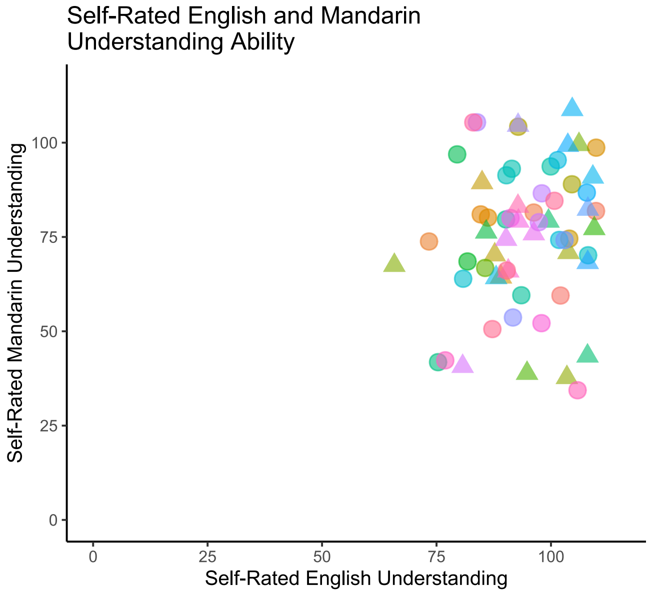
To determine if Mandarin understanding abilities are associated with individual differences in our participants’ perceptual sensitivity for ambiguity in a Singapore Mandarin alveolar-retroflex contrast continuum, we then conducted a one-tailed Pearson correlation analysis between our participants’ self-rated Mandarin understanding ability scores, and their individual decision gradient slope values. Individual Mandarin understanding ability scores were obtained directly from each participants' Language ID, and individual decision gradient slope values were extracted from the psychometric slopes fitted to their responses on the Spring-Village CROWN Game.

5.5 Results

5.5.1 Language ID

Figure 5.3

Participants' self-reported English and Mandarin understanding ability.



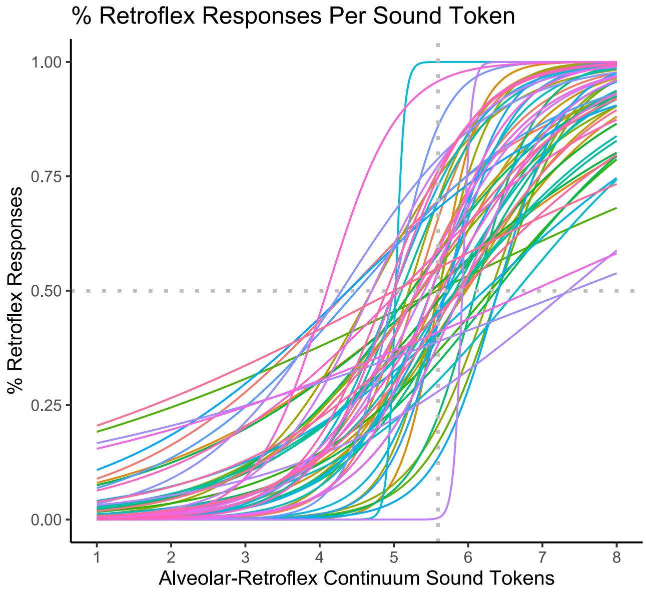
Note. $N = 62$. Circles: female, triangles: male. Responses jittered by 10 for visualisation.

Figure 5.3 shows the distribution of our participants' self-rated English and Mandarin understanding abilities. Overall, our participants self-reported high levels of understanding abilities for spoken English (Range: 60-100, *Median* = 95, *SD* = 8.2), indicating that they would not have had any issues in understanding the task instructions. Participants self-reported more varying levels of understanding abilities for spoken Mandarin (Range: 37-100, *Median* = 76, *SD* = 17.2).

5.5.2 Spring-Village CROWN GAME

Figure 5.4

Decision gradient slopes of individual participants, showing fitted psychometric functions.



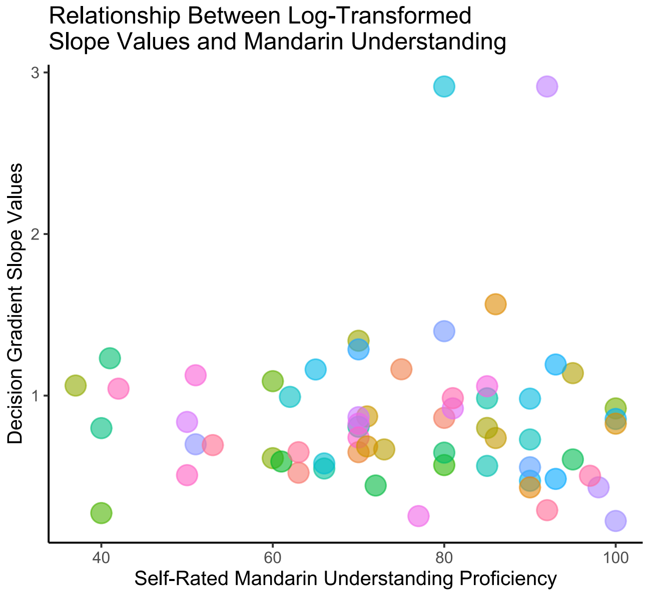
Note. *N* = 62. Dotted lines indicate median boundary threshold.

The individual decision gradient slope values of each participant are shown in Figure 5.4 together with the overall median alveolar-retroflex boundary threshold (represented by grey dotted lines). From the fitted psychometric functions, we obtained an overall median boundary threshold of 5.6 (*SD* = .67).

5.5.3 Link between Mandarin understanding and decision gradient slopes

Figure 5.5

Distribution of each participants' psychometric slope values and self-rated Mandarin understanding ability.



Note. *N* = 62. Mandarin understanding abilities rated on scale of 0 – 100.

Figure 5.5 shows each participant's decision gradient slope value plotted against their self-reported Mandarin understanding ability scores. As there were two slope value outliers in the dataset, we applied a log transformation to all slope values to approximate a Gaussian distribution before further analysis was conducted. The Pearson correlation analysis was then conducted on the log-transformed slope values to assess the strength of relationship between Mandarin understanding ability scores, and decision gradient slope values. The analysis revealed a weak correlation coefficient of $r = .02$, $t(1,60) = .14$; $p = .44$; $B_{H(0,\,0.25)} = 0.14$, indicating that there was no significant correlation between our participants' self-rated Mandarin understanding ability scores and their contrast sensitivity decision gradient slope values.

5.6 Discussion

Overall, the results of this study are consistent with that of our earlier word identification study. Our currently study shows that the majority of our Singapore Mandarin speaking participants demonstrate perceptual sensitivity response patterns consistent with the perception of a distinct Mandarin alveolar-retroflex phoneme contrast. Moreover, this study shows that this contrastive perception is not restricted to word identification alone, but can also be found with a fine-grained measure of perceptual sensitivity using the minimal word pair [tsʰuə́n]-[ʈʂʰuə́n]. As this minimal pair has been documented to contain higher degrees of phonological ambiguity (e.g., as shown in our earlier CoG analysis; Chang, 2012; Jeng, 2006), we can see that Singapore Mandarin speakers can demonstrate high levels of alveolar-retroflex phoneme contrast sensitivity for the local variety of Mandarin.

We also found a range of individual differences in our participants' patterns of ambiguity resolution across the Mandarin alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] continuum. A visual inspection of Figure 5.4 shows that some of our participants had visibly steeper decision gradient slopes, characteristic of high levels of sensitivity to different degrees of phonological ambiguity across the continuum. On the other hand, some participants had visibly shallower decision gradient slopes, characteristic of lower levels of perceptual sensitivity for phonological ambiguity across the continuum. This shows that fine-grained individual differences in phoneme contrast sensitivity can be effectively captured with this set of stimuli using the paradigm of the Spring-Village CROWN game.

Our examination of our participants' perceptual alveolar-retroflex boundary thresholds revealed that the median boundary threshold of our sample of Singapore Mandarin speakers lies between Sound 5 and Sound 6, closer to the retroflex end of the continuum. This indicates that on average, our participants showed a greater tolerance for ambiguity in alveolar articulations, with the majority of our participants judging a wider range of ambiguous sound tokens to be consistent with their representation of the voiceless *alveolar* affricate in [tsʰuə́n] as compared to the voiceless retroflex affricate in [ʈʂʰuə́n]. This boundary threshold patterns differs from existing documentation on how both Taiwan and Beijing Mandarin speakers resolve acoustic ambiguity on a Beijing Mandarin alveolar-retroflex contrast continuum. Chang (2012) found that the boundary thresholds of Taiwan and Beijing Mandarin speakers lie closer to the retroflex end of the continuum, indicating that unlike Singapore Mandarin speakers, they had a higher tolerance for ambiguity in retroflexion (with Taiwan Mandarin speakers showing more tolerance for retroflex ambiguity than Beijing Mandarin speakers). It is possible that this difference could be related to variations in the manner in which the alveolar and retroflex places of articulation are used in each of the different varieties of Mandarin. However, we cannot directly compare our results to that of Chang (2012), as the stimuli used in their study were synthesised from Beijing Mandarin, which is phonologically distinct from both Taiwan and Singapore Mandarin. Thus, the ambiguity resolution patterns of Chang's (2012) study apply specifically to how Taiwan and Beijing Mandarin speakers perceive the alveolar-retroflex contrast in the standard variety of Beijing Mandarin, and cannot be generalised to findings in other varieties of Mandarin.

Finally, we did not find evidence of any significant relationship between our participants' Mandarin understanding abilities and their perceptual sensitivity for ambiguity embedded in our Singapore Mandarin alveolar-retroflex contrast continuum. This shows that daily overall Mandarin speech comprehension does not appear to be related to the degree of sensitivity that Singapore Mandarin speakers for sensitivity to ambiguity along the alveolar-retroflex [tsʰuə́n] – [ʈʂʰuə́n] contrast. Thus, highly structured contrast sensitivity for the Mandarin alveolar-retroflex contrast does not appear to be needed for accurate speech comprehension in naturalistic daily Mandarin speech. This is consistent with an observational study of Taiwan Mandarin speakers which found that misplaced retroflex use does not appear to have an impact on overall speech comprehension (Chung, 2006). In fact, documentations of naturalistic conversational speech in Taiwan Mandarin have shown that Taiwan Mandarin speakers may deliberately utilise deretroflexion or hypercorrection (where retroflexion is used in place of the alveolar place of articulation) as a stylistic speech choice to highlight concepts and enhance listening comprehension in daily conversation (Chung, 2006).

Similarly, Singapore Mandarin speakers may have a more flexible understanding of how different places of articulation should be used in Mandarin, and may instead rely more heavily on other naturalistic speech cues (e.g., context) to distinguish between words. This could be a more efficient speech comprehension strategy in daily conversation, as Mandarin contains a large number of homophones that cannot be disambiguated through strict phoneme contrast identification alone.

5.7 Conclusion

On the whole, our fine-grained perceptual analysis showed that Singapore Mandarin speakers do in fact perceive a distinct contrast in an alveolar-retroflex contrast acoustically aligned to the norms of the local variety of Mandarin. We have also found that a range of individual differences can be found in our participants' perceptual sensitivity for different degrees of ambiguity embedded in our continuum of voiceless alveolar-retroflex affricate [tsʰuə́n] – [ʈʂʰuə́n] sound tokens. While some participants display high levels of ambiguity resolution sensitivity, others display less sensitivity for different degrees of alveolar-retroflex ambiguity. However, these individual differences in perceptual sensitivity do not appears to be related to our participants' ability to comprehend spoken Mandarin in daily speech. This suggests that Singapore Mandarin speakers may not require high levels of structured phoneme contrast sensitivity to have functional levels of conversational Mandarin comprehension.

Our findings have helped to establish a set of stimuli that is valid for the local population of contemporary Singapore Mandarin speakers. The use of such validated stimuli will ensure that future studies are well-suited to the speech norms of Singapore Mandarin speakers, and can be used to inform educators on how local students perceive and understand speech sounds. We have also made our stimuli and Spring-Village CROWN game task available in an online repository (https://osf.io/rnq4w) to facilitate replications and extensions of our research. With this study, we have established that a fine-grained measure of perceptual sensitivity for the Singapore Mandarin alveolar-retroflex contrast can be obtained with our Spring-Village CROWN game, we are confident that this this set of stimuli and this paradigm can be used to effectively assess different aspects of how Singapore Mandarin speakers perceive this Mandarin alveolar-retroflex contrast. We now have a validated method of assessing perceptual sensitivity for the Singapore Mandarin alveolar-retroflex contrast with a novel set of stimuli aligned to the acoustic norms of contemporary Singapore Mandarin. Thus, we are confident that we will be able to investigate a language-specific transfer effect in Singapore English-Mandarin bilinguals with a distributional learning task.

Before we began our distributional learning study, we first conducted an exploratory study to examine any possible links between perception and production of the Singapore Mandarin alveolar-retroflex contrast. Existing studies have demonstrated that there is a link between the degree of contrast in perception and production of phonemes by native language speakers (Fox, 1982; Newman, 2003). Moreover, evidence also suggests that perceptual sensitivity training on phoneme contrasts can lead to corresponding increases in contrastiveness of phoneme contrast in speech production (Inceoglu, 2016; Sakai & Moorman, 2018). We thus conducted a small-scale exploratory study to investigate if any such links exist between perception and production of the Singapore Mandarin alveolar-retroflex contrast. Based on the results that we find, we may design future studies to determine if distributional training for perceptual sensitivity to our Singapore Mandarin [tsʰuə́n]-[ʈʂʰuə́n] alveolar-retroflex contrast might lead to changes in contrastiveness in speech production.

Chapter 6: Investigating speech perception-production links for a Singapore Mandarin alveolar-retroflex sibilant contrast

6.1 Introduction

In Chapter 4, we examined the fine-grained acoustic characteristics of speech production of the Singapore Mandarin alveolar-retroflex contrast, and found that Singapore Mandarin speakers have a range of individual differences in the acoustic characteristics of their production of the two categories of speech sounds. We also conducted a fine-grained examination of how Singapore Mandarin speakers perceive an alveolar-retroflex contrast in Singapore Mandarin in Chapter 5, and found a wide range of perceptual patterns in how Singapore Mandarin speakers resolve acoustic ambiguity for speech sounds in this contrast. As studies have suggested that there may be links between perception and production of phoneme contrasts, we set out to examine possible links between individual differences in how Singapore Mandarin speakers perceive and produce the Singapore Mandarin alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast. Although we have previously collected measures of Singapore Mandarin alveolar-retroflex speech production and perceptual word recognition in Chapter 4, the results we obtained were not well suited for an assessment of links between speech perception and production. We were unable to clearly assess individual differences in perception in Chapter 4 as the majority of the participants performed at ceiling level accuracy for the word identification task. Moreover, we were unable to obtain suitably high quality audio recordings of speech production in Chapter 4 as the study took place at the height of the COVID-19 pandemic, we were not able to access a high quality sound-attenuated recording environment due to social distancing restrictions. As COVID-19 restrictions have since been relaxed, we were able to conduct speech recordings for Chapter 6 in an on-campus sound-attenuated recording booth. This allowed us to obtain high quality audio recordings without high ambient noise levels, making this data well suited for examining fine-grained links between frication perception (with the Spring-Village CROWN Game) and production (with the speech elicitation task in Chapter 4). As this study was entirely exploratory in nature, we did not conduct a preregistration prior to data collection as we did not have any particular expectations for the pattern of results we would find. However, in line with open science principles, all our task files, analysis files, and the data analysed in this study have been saved in a repository on the Open Science Framework and are available here https://osf.io/nh9sw/.

6.1.1 Theories and evidence on auditory-motor links in speech perception

Theories on the manner in which motor activation in speech production is involved in the perception of speech signals have been extensively discussed in the existing literature. These theories suggest that motor representations of speech production play an integral role in the perceptual processing of speech. One of these classes of theories – Motor Theories – suggest that speech perception extends beyond the auditory domain, with links existing between the perception and production of speech. According to Motor Theory (Lieberman & Mattingly, 1985), speech perception is achieved through directly referencing how speech is produced by the listener. This is achieved by matching the acoustics of the speaker voice to the listener's own model of how they would move their own articulators to produce the same speech sounds. The Analysis by Synthesis theory (Stevens & Halle, 1967) further suggests that while motor gestures are involved in speech perception, the exact process is less direct and occurs as a result of listeners mentally synthesising incoming speech signals into corresponding motor commands – these synthesised motor commands are then referenced against the perceived speech signals to accurately decipher the incoming information.

Indeed, there is evidence of motor regions of the brain being involved in speech perception. Such studies have been conducted with both adults and infants, suggesting that the generation of motor models may indeed be involved in the perception of speech. Callan et al. (2014) conducted an fMRI study to test native English speakers and Japanese speaker on their discrimination of the English /r/-/l/ contrast. The results of their study revealed different patterns of auditory and motor region activation based on whether or not the contrast was present in the participants' native language. In particular, English speakers in their study showed greater activation in the auditory areas such as the superior temporal gyrus (STG) as compared to motor areas of the brain, while the Japanese speakers showed the inverse, with less auditory region activation, and greater activation of motor regions such as Broca's area, the premotor cortex, and the anterior insular (Callan et al., 2014). Similarly, Kuhl et al. (2014) conducted a series of MEG studies to compare regions of brain activation in 7-

month-old and 11-to-12-month-old infants when they were presented with native vs. non-native phoneme contrasts. As discussed in Section 1.3.1, infants typically begin life sensitive to a wide variety of speech sounds, and begin losing sensitivity for non-native phoneme contrasts at around 10 months of age. Kuhl et al. (2014) found that the 7-month-olds who had not yet developed native-language phoneme perceptual sensitivities showed equal amounts of activation in the STG and Broca's area. On the other hand, the 11-to-12-months olds who had already begun developing sensitivities to native language phoneme contrasts showed less activation in Broca's area as compared to the STG. The results of Callan et al. (2014) and Kuhl (2014) have been suggested by the authors to be indicative of auditory-motor links which become innate for native speech sounds due to extensive listening experience, but which require higher levels of effortful activation for the construction of mental motor models of non-native speech sounds. The patterns of brain region activation observed in these studies appear to suggest that the motor activation of speech production is likely to be important for the perception of speech. More directly, studies also show that speech perception-production links can be found in behavioural studies of phoneme perception and production.

6.1.2 Behavioural evidence for perception-production links in speech

Behavioural studies have shown that there appear to be links between speech perception and production of speech across the lifespan. Infant studies have demonstrated that such a link can be observed from very early on in life. For instance, Kuhl and Meltzoff (1982) demonstrated that 18- to 20-month old infants were able to accurately match videos of lip movements to audio recordings of speech sounds, suggesting that the ability to map articulatory motor gestures onto speech sounds is present from early life. Tsao et al. (2004) also found a link between speech perception in even younger infants and speech production abilities later in infancy by demonstrating that phonetic discrimination sensitivities in 6-month-olds were correlated with their language and vocabulary development later in life at 24 months of age.

More direct links between perception and production of speech can also be found in studies of children with specific articulatory disorders. These studies typically show that children with articulatory disorders specific to certain phonemes tend to demonstrate corresponding difficulties in perceiving the same phonemes in speech. For instance, Rvachew & Jamieson (1989) tested 5- and 7-year-old children with functional articulatory disorders on their perceptual sensitivity to different phoneme contrasts and found that the children had perceptual sensitivity difficulties for the phoneme contrasts their articulatory disorders were linked to. A large subset of the 5 year-old children with articulatory disorders linked to the voiceless frication /ʃ/ showed difficulties at accurately identifying words with a /s/-/ʃ/ initial place of articulation. At the same time, all the 7-year-olds with articulatory disorders specific to the voiceless frication phonemes /s/ and /θ/ demonstrated similar difficulties at identifying words with a /s/-/θ/ initial place of articulation. The results of this study appear to show that functional articulatory disorders in children are linked to perceptual difficulties specific to the phonemes that they have difficulty in producing.

Similarly, Hearnshaw et al. (2018) investigated differences in speech perception-production links in Australian English speaking children with and without speech sound disorders (SSD) using word identification and speech production tasks for words with the initial phonemes of /k/, /ɹ/, /ʃ/, and /s/. While the study found that children with Australian SSDs appeared to show generalised difficulties in perceiving words with all four initial phonemes as compared to the children with typical speech development, the study also revealed an overall correlation in speech perception and production accuracy for all participants. Moreover, Hearnshaw et al. (2018) also found an overall trend of all participants showing more difficulties in producing /ɹ/-initial words. As Australian English is a non-rhotic variety of English, it is possible that this could be linked to the participants' lack of perceptual experience with the /ɹ/ phoneme in daily conversation. Therefore, these studies show that difficulties in the production of specific speech sounds may be linked to corresponding difficulties in perception the same categories of speech sounds.

At the same time, studies on adults have also demonstrated that improvements in perception may be linked to corresponding improvements in production of certain phonemes. Studies focusing on non-native phoneme contrast training have found that training in the perceptual domain often has effects that carry over to the production domain of speech. For instance, Bradlow et al. (1997) conducted a speech perception and production study on Japanese-speaking adults, in which they trained participants on their perception of a novel English /r/-/l/ contrast. The phonemes /r/-/l/ are non-contrastive in Japanese, and are therefore difficult for native Japanese speaking adults to perceive and produce. Bradlow et al. (1997) found that participants who showed improvements in perceptual sensitivity for the contrast following training were more likely to also showed concomitant improvements in their production of the contrast. Similarly, Inceoglu (2016) trained English speakers on their perceptual sensitivity for the French /ɔ̃/, /ɑ̃/, /ɛ̃/ vowels which are not present in English. Inceoglu (2016) likewise found that their participants' improvements in perceptual sensitivity for the vowel contrasts corresponded with improvements in production of the vowels. Moreover, this pattern of results is borne out in a meta-analysis conducted by Sakai and Moorman (2018) on 30 studies of novel phoneme perception training followed by speech production assessment. This meta-analysis revealed that perceptual training for novel phoneme contrasts generally results in at least a small to medium improvement in speech production that is stable over time (Sakai & Moorman, 2018). Thus, evidence appears to show that links exists between improvements in perceptual sensitivity to novel phonemes, and corresponding improvements in production of said phonemes.

Studies examining the production and perception of phonemes in native language speakers have also shown evidence for perception-production links. For instance, Fox (1982) tested English speaking participants on how constrastively they perceived and produced English vowels in terms of acoustic space. The results of this study revealed a significant correlation between the two, with participants who reported perceiving the vowels more contrastively from each other also producing vowels that were at a greater distance from each other in acoustic space (Fox, 1982). Similarly, Newman (2003) assessed native English speakers for a link between contrastiveness in perception and production of the English consonant contrasts /b/-/p/ or /s/-/ʃ/. Newman (2003) obtained data on participants' speech production in a task in which participants were recorded producing the syllables /ba/ and /pa/ or /sæ/ and/ʃæ/. Following this, information was obtained on participants' perceptual prototypes of each phoneme using a “goodness of fit” rating scale in which participants were presented with sound tokens on a /ba/-/pa/ or /sæ/-/ʃæ/ continuum and told to rate each token on how well it fit the category of /p/ or /ʃ/. The results of Newman's (2003) study revealed that there was a significant link between participants' perception and production of the phonemes, with the participants giving higher ratings to speech sounds that were more closely aligned to their own productions of the phonemes in terms of voice onset time (VOT) and peak spectral frequency. From these studies, we can see that there is evidence that differences in the how phoneme contrasts are perceived in acoustic space are correspondingly linked to the acoustic characteristics of the same phonemes when produced in speech.

On the other hand, some studies appear to show weaker links between perception and production of speech in native language speakers. For instance, Frieda et al. (2000) conducted a study on the link between hyperarticulation in perception and production of the English vowel /i/ - this study proposed that there would be a link between perceptual preferences for hyperarticulation, and hyperarticulation of the vowel in speech. To obtain data on participants' perceptual prototypes of the vowel /i/ participants were instructed to select their perceived best exemplar tokens out of 330 /i/ tokens varying incrementally in F1 and F2. Following this, participants were instructed to produce speech tokens of words with the vowel /i/ in naturalistic speech, or in "clear" speech to approximate hyperarticulation. F1 and F2 data were then extracted from the speech recordings tokens, and compared to the perceptual tokens. However, the link that Frieda et al. (2000) expected to find was not fully supported by the results of the study. While a small subset of participants did indeed have a link between their chosen /i/ vowel prototypes and the degree to which they hyperarticulated /i/ in speech, the majority of participants showed a pattern of preferring vowel prototypes with hyperarticulation that exceeded even that of their own "clear" speech production tokens. Similarly, a study by Bailey and Haggard (1980) examined links between perception and production in 3-year-old children, but did not find a significant positive correlation between their averaged perceptual discrimination boundaries across a series of voiced-voiceless /b/-/p/, /d/-/t/, /g/-/k/ VOT continuums, and the averaged VOTs of their productions of different minimal pairs of words with an initial voiced/voiceless consonant (i.e., bin/pin, bear/pear, deer/tear, goat/coat, girl/curl). While Bailey and Haggard (1980) *did* report finding a moderate correlation between the averaged production boundary threshold and the perceptual boundary threshold for /b/-/p/, no similar links were found for /d/-/t/ or /g/-/k/. Thus, these studies demonstrate that perception-production links in speech may vary in strength between individuals and conditions.

It is possible, however, that methodological differences could have contributed to the contrasting results found in the studies discussed above. For instance, it is possible that requiring participants to identify best-fit examples of vowels (as in Frieda et al. 2000) could have resulted in participants choosing perceptual prototypes with greater degrees of hyperarticulation than in naturalistic speech. This could occur as a result of the Hyperspace Effect proposed by Johnson et al. (1993). According to this Hyperspace Effect, native language speakers typically store perceptual templates of phonemes in more hyperarticulated forms, leading to highly hyperarticulated tokens being preferred in tasks of best-fit phoneme token selection (Johnson et al, 1993). At the same time, it is worth noting that the speech tokens used in the perception and production analysis of Bailey and Haggard’s (1980) study were not always directly matched, collapsing across different places of articulation and using different minimal pairs of words to obtain overall measures of perception and production. This could have impacted the results of their study, as collapsing across categories of speech sounds and minimal pairs of words fails to account for individual sources of variance from different places of articulation, and different coarticulation effects from vowels across different minimal pairs of words. Moreover, the participants in Bailey & Haggard’s (1980) study consisted of 3-year-old children who were assessed on their perception of word tokens synthesised to the norms of adult speech. This could have led to an additional source of mismatch in the stimuli. Studies have demonstrated that there are significant differences between both perception and production of voiced-voiceless phoneme contrasts in adults and children. For instance, Zlatin and Koenigsknecht (1976) compared the perception and production of similar voiced-voiceless initial consonant minimal pairs of words in 2- and 6-year-old children and adults. This study revealed that there are significant differences between the perception and production of young children and adults, with adults typically producing speech tokens more contrastively, and having a narrower phoneme contrast boundary in perception as compared to children. This discrepancy could therefore have introduced further complications to the results of Bailey & Haggard's (1980) study. Thus, any perception-production links that do exist in speech could have been attenuated.

6.1.3 Analysis of perception-production link in Singapore Mandarin alveolar-retroflex contrast

Based on what we have found in existing literature, we designed our investigation of a perception-production link in Singapore Mandarin to ensure that the target sound tokens used in perception and production were as closely linked as possible. For our measure of speech perception, we chose to use our synthesised 8-step Singapore Mandarin 村 [tsʰuə́n] (village) - 春 [ʈʂʰuə́n] (spring season) affricate alveolar-retroflex continuum that we created in

Chapter 5 based on the acoustic norms of Singapore Mandarin speakers in Chapter 4. For our measures of speech production, we obtained spoken word recordings of the words [tsʰuə́n] and [ʈʂʰuə́n] from native speakers of Singapore Mandarin. Finally, in order to ensure that our measure of speech perception would not be affected by a task-dependent hyperspace effect (Johnson et al. 1993), we used our fine-grained measure of alveolar-retroflex contrast perception – the Spring-Village CROWN Game used in Chapter 5. This paradigm would allow us to obtain fine-grained measures of phoneme contrast sensitivity across a continuum without requiring participants to choose a best-fit prototype of each sound.

While there does not yet exist evidence pointing towards a perception-production link for Mandarin alveolar-retroflex contrasts, some studies appear to show that there are links between the variety of Mandarin that is spoken and the manner in which the standard Beijing Mandarin alveolar-retroflex contrast is perceived. For instance, Chang (2012) and Chang et al. (2013) investigated Taiwan and Beijing Mandarin speakers’ perception of an alveolar-retroflex contrast continuum in standard Beijing Mandarin, and found that Taiwan Mandarin speakers were typically more accepting of a wider range of acoustic ambiguity in retroflexion as compared to standard Beijing Mandarin speakers. Chang et al. (2013) suggested that this pattern of perceptual responses reflects how clearly retroflexion is produced in Taiwan Mandarin (less distinct from the alveolar to deretroflexion) as compared to Beijing Mandarin (more distinct from the alveolar). This suggests that we might likewise be able to obtain evidence of perception-production links in a Singapore Mandarin alveolar-retroflex contrast by examining individual difference in our participants' Mandarin speech perception and production.

To promote open science practices, we have made all relevant data, task files, stimuli, and analysis files for this study available on the Open Science Framework (https://osf.io/nh9sw/).

6.1.3.1 Acoustic analysis of production of Singapore Mandarin [tsʰuə́n]-[ʈʂʰuə́n] contrast

To obtain speech production data from our participants, we conducted the Spring-Village speech production task first used in Chapter 4, in which word tokens were recorded from participants reading out loud from a wordlist of alveolar-retroflex minimal pairs of words (see Table 3.1 for wordlist). However, we focused our acoustic analyses on the word tokens 村 [tsʰuə́n] (village) - 春 [ʈʂʰuə́n] (spring season) only. As in our earlier analysis of the Spring-Village speech production task in Chapter 4, we focused on two main analyses to first examine differences between productions of frication for the alveolar [tsʰuə́n] and retroflex [ʈʂʰuə́n] in Singapore Mandarin. We conducted an analysis on the Centre of Gravity (CoG) values of each frication token, and a general additive mixed model analysis (GAMM) on the long-term averaged spectrum (LTAS) data of each frication token.

For the examination of perception-production links in speech, we also obtained measures of alveolar-retroflex contrastiveness in speech from the CoG values and LTAS matrices. A measure of alveolar-retroflex CoG contrastiveness (CoG Δ) was obtained for each participant from the difference between the mean of each participant's [tsʰuə́n] and [ʈʂʰuə́n] frication CoG values. Larger CoG Δ values represent a greater degree of contrast in speech and smaller CoG Δ values represent less contrast in speech. A measure of alveolar-retroflex contrastiveness was also obtained for each participant from their LTAS matrices.

6.1.3.2 Analysis of perceptual sensitivity for Singapore Mandarin [tsʰuə́n]-[ʈʂʰuə́n] contrast

To obtain fine-grained measures of alveolar-retroflex contrast perceptual sensitivity, we used the Spring-Village CROWN game (first used in Chapter 5) to assess participants' patterns in ambiguity resolution for an 8-step [tsʰuə́n]-[ʈʂʰuə́n] Singapore Mandarin continuum. As in our earlier analysis of the Spring-Village CROWN game in Chapter 5, we obtained individual boundary threshold values for each participant to determine the point at which their sound token identification decisions switched from the alveolar-initial word [tsʰuə́n] to the retroflex-initial word [ʈʂʰuə́n]. We also obtained individual decision gradient slope values as a measure of alveolar-retroflex contrastiveness in perception, with steeper slopes and higher decision gradient slope values representing more contrastive speech perception.

6.1.3.3 Perception-production link for Singapore Mandarin [tsʰuə́n]-[ʈʂʰuə́n] contrast

For our investigation of perception-production links in the alveolar-retroflex contrast in Singapore Mandarin, we focused on three main links of interest. First, we examined for links between contrastiveness in perception and production of the alveolar-retroflex frication contrast. We assessed for relationships between perceptual decision gradient slope values and CoG Δ values, and perceptual decision gradient slope values and alveolar-retroflex difference LTAS matrices.

Next, we examined for links between perceptual alveolar-retroflex boundary thresholds, and productions of Mandarin alveolar and retroflex frication. We assessed for relationships between perceptual boundary thresholds and individual alveolar and retroflex CoG values. We also assessed for relationships between perceptual boundary thresholds and individual alveolar and retroflex LTAS data.

6.2 Speech Production: Spring-Village Speech Production Task

To ensure that there would be no accent transfer effects from the perception task to the speech production task, all participants completed the speech production task before completing the perception task. This speech production task was a methodological replication of the Mandarin alveolar-retroflex Spring-Village speech production task in Chapter 4.

6.2.1 Methods

6.2.1.1 Participants

21 participants (15 female, 5 male, and 1 participant who identified as non-binary on the perception task, and female on the production task) were recruited via an NTU student studies recruitment group in exchange for $5 for every half hour of their participation. One male participant was excluded from the study due to not completing all the tasks. The remaining sample of 20 participants were aged between 20 – 33 years of age (Median = 23.5, *SD* = 3.83). All participants self-identified as English-Mandarin bilinguals, and reported that they had grown up in Singapore and had not spent more than one year living outside of Singapore. Informed consent was collected from all participants prior to commencement of the study, and this study was approved by the IRB of Nanyang Technological University institution (IRB-2019-01-034).

6.2.1.2 Equipment

All stimuli were presented on a 10-inch LCD Microsoft Surface Go. Speech recordings were conducted in a sound-attenuated recording booth. Participants spoke into a microphone with a pop filter, and speech was recorded with a Zoom H4n Pro Handy Recorder. All recordings were made in stereo WAV format, with a sample rate of 44.1kHz and a bit depth of 16 bits.

6.2.1.3 Stimuli

Visual stimuli for the speech production task were identical to that of the Mandarin alveolar-retroflex Spring-Village speech production task of Chapter 4, and consisted of 12 Mandarin word picture cards. The 12 words comprised six alveolar-retroflex minimal pairs. Table 3.1 contains the full list of minimal pairs, and Figure 4.5 shows an example of a picture card. Each card featured an illustrated depiction of the meaning of the word, along with the Chinese character of the word, and its corresponding pronunciation in Hanyu Pinyin. The full set of illustrations and word picture cards can be found in our online repository (https://doi.org/10.21979/N9/ZTMPML).

No auditory stimuli were used for this task.

6.2.2 Procedure

The procedure for the speech production task was identical to that of the Mandarin alveolar-retroflex Spring-Village speech production task of Chapter 4.

6.2.2.1 Spring-Village speech production task

Participants were first introduced to the speech production task with written instructions in English. Participants were informed that a picture card would appear on the screen, and their task was to read each word aloud three times with a three second pause between each reading. To ensure that speech was produced as naturalistically as possible, participants were instructed to pronounce the words in a manner similar to that of daily speech. Emphasis was placed on the fact that the task was not a test of pronunciation accuracy, and that we were simply interested in documenting the unique speech patterns of Singapore Mandarin speakers.

Following the instructions, participants completed one block of audio recording consisting of 12 trials. For each trial, participants were presented with one of the 12 Mandarin alveolar-retroflex picture cards together with written instructions in English prompting them to read each word three times. Figure 4.5 shows an example of the visual stimuli presented in a trial. After reading each word three times, the participant made a keypress to move on to the next trial. The 12 picture cards were presented to each participant in random order. No feedback was given to the participants so as to not influence their speech production. No trial timeout was implemented to ensure that participants had sufficient time to read all the words presented.

6.2.2.2 Language ID

To collect data on participants' language understanding abilities, they were asked to rate how well they understood the four main languages in Singapore (English, Mandarin, Malay, and Tamil) on a scale of 0 to 100 using an onscreen sliding scale. Participants were also given an option to add in more languages or dialects to rate.

6.2.3 Analysis Plan

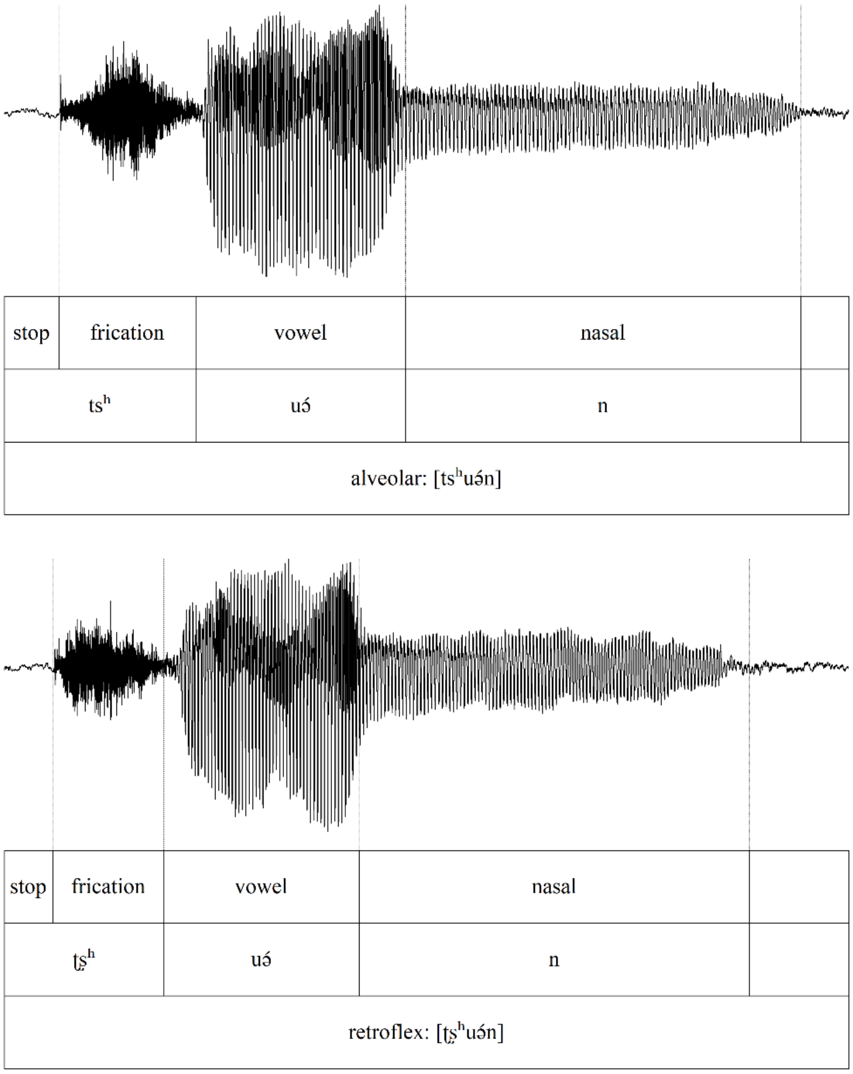
6.2.3.1 Audio processing

The audio processing was identical to the audio processing conducted for the Mandarin alveolar-retroflex Spring-Village speech production task in Chapter 4 with the exception that frication segmentation was only conducted on [tsʰuə́n] and [ʈʂʰuə́n] word tokens. Individual word tokens were obtained from each of the participants' audio files. Each word was manually selected, labelled, and spliced from the raw audio file in Goldwave [v.6.51], and saved as separate audio files in WAV format. The complete set of raw audio files and individual word tokens are archived in the Spring Village Corpus under the Growing Collection of audio recordings (Goh et al., 2022) and can be accessed here: https://doi.org/10.21979/N9/ZTMPML.

Following this, the frication portion of each segmented [tsʰuə́n] and [ʈʂʰuə́n] token was manually selected based on visual inspection of the waveform, labelled, and spliced in Praat 6.1.16 (Boersma & Weenink, 2020). Each individual frication token was saved as separate files in short txt and wav format. Figure 6.1 shows the identified frication portions of an alveolar word [tsʰuə́n] and a retroflex word [ʈʂʰuə́n] token produced by one participant (Participant 7). Figure 6.2 shows the spectral slices of the full spectra of the extracted frication of an alveolar word [tsʰuə́n] and a retroflex word [ʈʂʰuə́n] token produced by one participant (Participant 7).

Figure 6.1

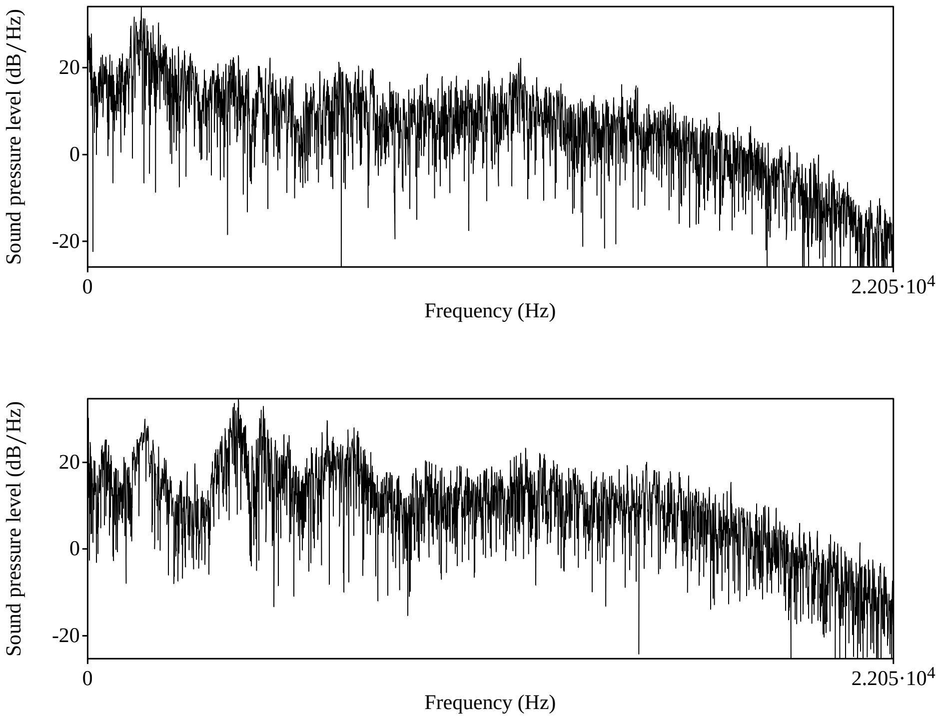
Waveform of alveolar [tsʰuə́n] (top) and retroflex [ʈʂʰuə́n] (bottom).



Note. Both recorded from the same participant (Participant 7). Segmentation is conservative such that no voicing from the following vowel is included in the voiceless segment for analysis.

Figure 6.2

Spectral slices of frication spectrum of alveolar [tsʰuə́n] (top) and retroflex [ʈʂʰuén] (bottom).



Note. Both recorded from the same participant (Participant 7).

6.2.3.2 CoG analysis

Following the methodology of Chang and Shih (2015) and Tso (2017), frication CoG values were obtained from the middle 30ms of the frication spectra of each token using the centre of gravity function in Praat 6.1.16 (Boersma & Weenink, 2020). Summary statistics were obtained for the alveolar [tsʰuə́n] and retroflex [ʈʂʰuén] CoG values obtained in this speech production task.

To determine if there was a significant difference between the CoG values of the frication of the alveolar [tsʰuə́n] and retroflex [ʈʂʰuén] produced by our Singapore Mandarin speakers, a linear mixed effects model analysis was conducted on the CoG data in R (R Core Team, 2020) with the *lme* function from the *nlme* package (Pinheiro et al. 2022). As we had a very small number of male participants in our sample (*N* = 4), we collapsed across gender for all analyses. The CoG analysis therefore looked at the categorical fixed factor of place of articulation (alveolar vs retroflex), with random intercept for participant and random by-participant slope for test phase.

6.2.3.3 LTAS GAMM analysis

To conduct the GAMM analysis, we exported the full frication spectra of each word token as LTAS matrices from Praat 6.1.16 16 (Boersma & Weenink, 2020). Using the *bam* function from the *mgcv* package (Wood, 2011) in R (R Core Team, 2020), we conducted a GAMM analysis on the LTAS matrices with the parametric fixed term of place of articulation (alveolar vs retroflex), smooth term of place of articulation, and random smooths for participant and trial number. Based on Sóskuthy (2017), a significant effect of a parametric term indicates a significant difference in the overall spectral density of the modelled data for each condition of the parametric term. Thus, we reported the effects of each parametric term assessed in the model. To assess for differences within the modelled spectral trajectories of each place of articulation, individual windows of spectral difference between the alveolar and retroflex places of articulation were identified using the *plot_diff* function from the *itsadug* package (van Rij et al., 2020) in R (R Core Team, 2020).

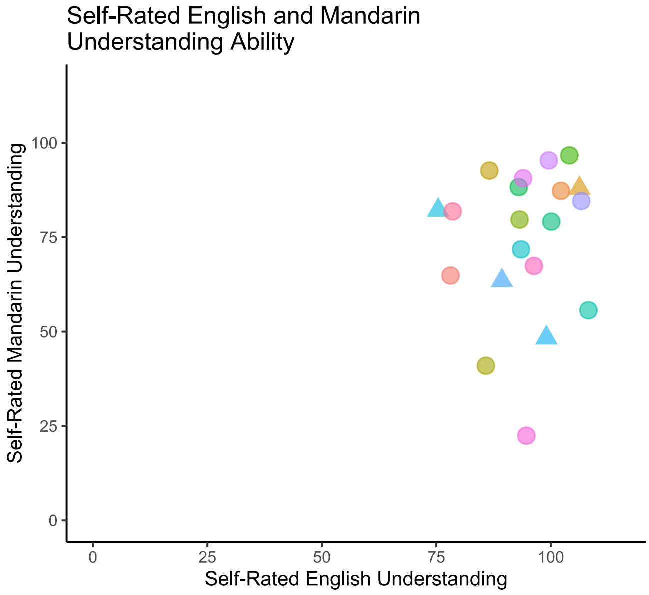
To ensure that the basis dimensions of the GAMMs would not result in oversmoothing of the fitted data, we conducted model criticism checks for each GAMM analysis using the *gam.check* function of the *mgcv* package (Wood, 2011). Based on the criticism checks comparing the EDF values of the model, we selected a *k*-value of 70 for this GAMM analyses to ensure that sufficient basis functions were included in the model to avoid oversmoothing of the data.

6.2.4 Results

6.2.4.1 Language ID

Figure 6.3

Participants' self-reported English and Mandarin understanding ability.



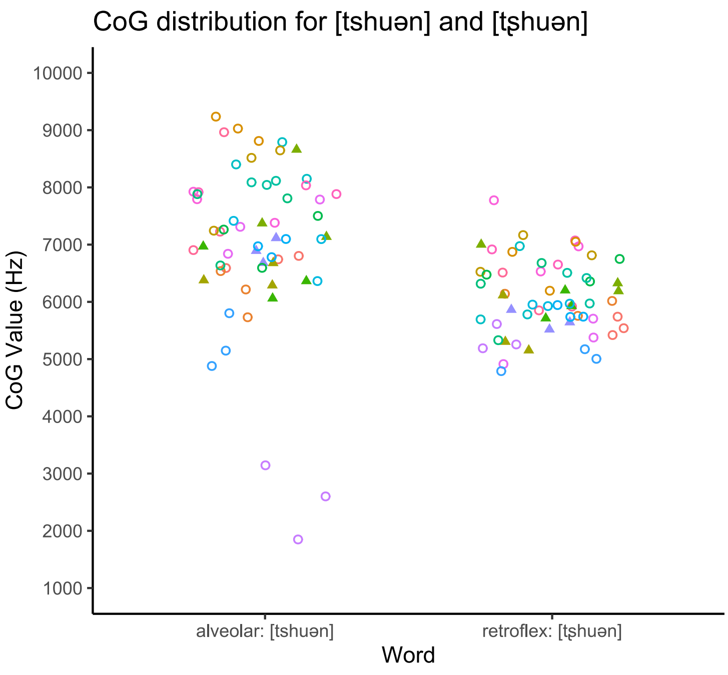
Note. *N* = 20. Circles: female; triangles: male. Responses jittered by 10 for visualisation.

Figure 6.3 shows the distribution of our participants' self-rated English and Mandarin understanding abilities. Overall, our participants self-reported high levels of understanding abilities for spoken English (Range: 70-100, *Median* = 100, *SD* = 9.2), indicating that they would not have had any issues in understanding the task instructions. Participants self-reported more varying levels of understanding abilities for spoken Mandarin (Range: 20-100, *Median* = 81, *SD* = 19.0). While one of our participants reported a Mandarin understanding ability that was much lower than our other participants (score of 20), all participants had taken part in formal education in Singapore and would likely have received at least 10 years of formal education in Mandarin. Moreover, the words that we chose for our Mandarin speech production and perception tasks were easily understandable by young children. Therefore, we did not have reason to exclude this participant from our analyses.

6.2.4.2 CoG results

Figure 6.4

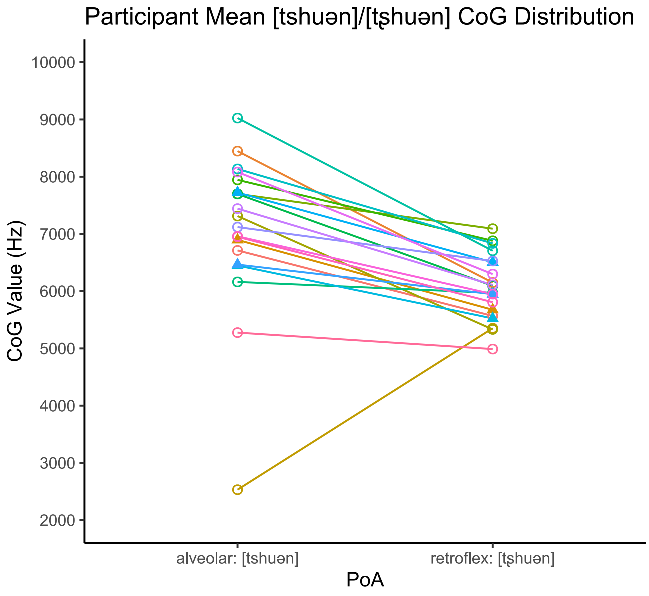
Distribution of each participant's alveolar [tsʰuə́n] and retroflex [ʈʂʰuén] CoG values for each recorded word token.



Note. *N* = 20. circles: female, triangles: male.

Figure 6.5

Mean CoG values of each participant by gender and PoA.



Note. *N* = 20. Circles: female, triangles: male.

Table 6.1

Summary statistics measured for [tsʰuə́n] and [ʈʂʰuén] CoG values.

	Alveolar [tsʰuə́n] CoGs (Hz)		Retroflex [ʈʂʰuén] CoGs (Hz)	
	Mean (SD)	Range	Mean	Range
Overall	7051.5 (1408.9)	1849.5-9235.99	6065.6 (642.2)	4789.9-7774.9
Female	7093.6 (1540.5)	1849.5-9236.0	6103.7 (671.8)	4789.9-7774.9
Male	6883.0 (684.9)	6060.6-8660.2	5912.9 (502.2)	5153.3-7001.0

Note. Standard deviation in parentheses.

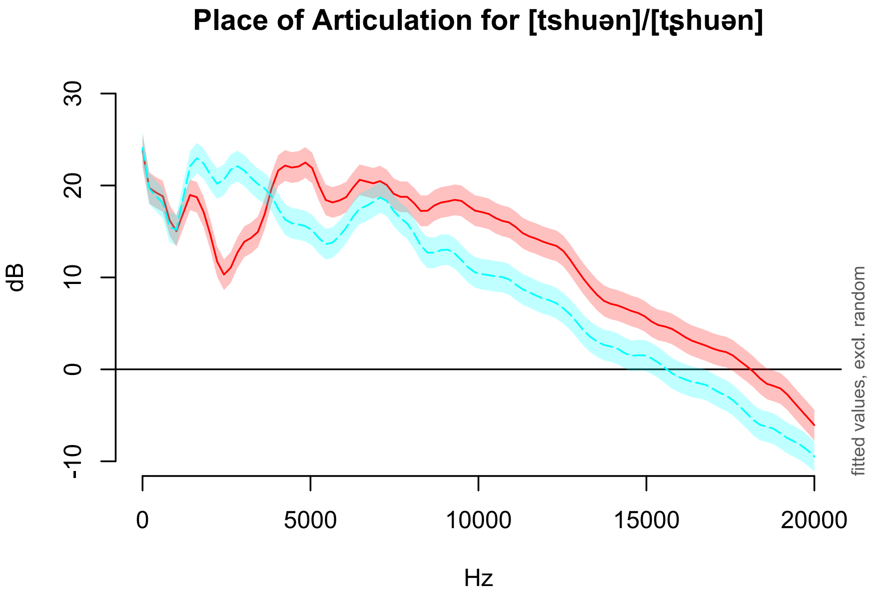
Figure 6.4 shows the CoG values obtained from each individual alveolar [tsʰuə́n] and retroflex [ʈʂʰuén] word token. Figure 6.5 shows the difference between mean alveolar [tsʰuə́n] and retroflex [ʈʂʰuén] CoG values of each participant. Overall summary statistics for the CoG values are presented in Table 6.1.

The linear mixed effects model analysis conducted on the CoG values revealed that there was a significant main effect of place of articulation, $t(1,59) = -6.58$, $p < .000001$, $\eta p^2 = .42$, indicating that there was a significant difference between our participants' overall CoG values for their alveolar [tsʰuə́n] and retroflex [ʈʂʰuén] frication, with higher CoG values for the alveolar [tsʰuə́n] tokens, and lower CoG values for retroflex [ʈʂʰuén] tokens (see Table 6.1 for summary statistics).

6.2.4.3 GAMM LTAS results

Figure 6.6

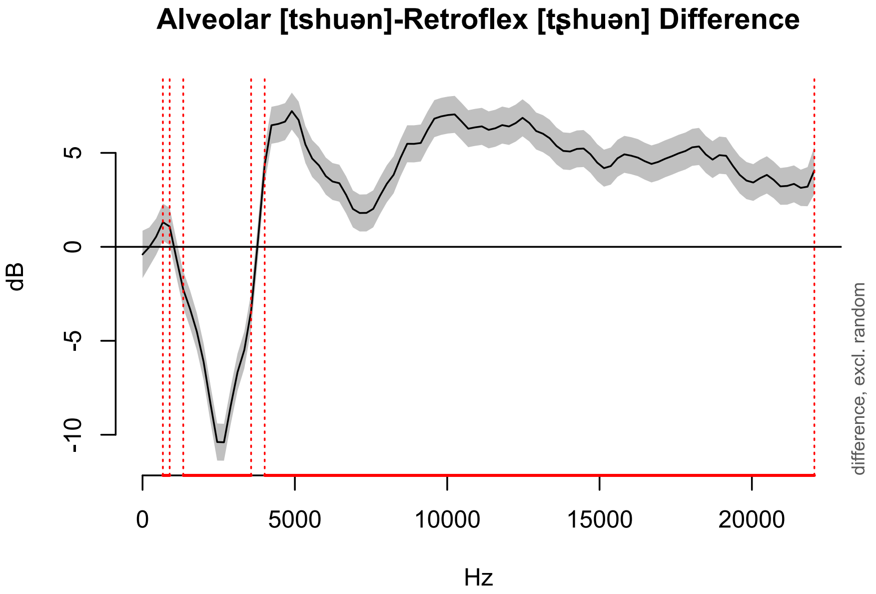
Fitted models of alveolar and retroflex frication LTAS data.



Note. Alveolar: red line; retroflex: turquoise dotted line.

Figure 6.7

Model of spectral difference between alveolar and retroflex frication LTAS data.



Note. Windows of significant difference bounded in red dotted lines.

The overall fitted LTAS models of our participants' alveolar and retroflex frication are presented in Figure 6.6, plotted with the *plot_smooth* function from the *itsadug* package (van Rij et al., 2020). The individual frequency windows of significant difference between the alveolar and retroflex fitted models are presented in Figure 6.7, plotted with the *plot_diff* function from the *itsadug* package (van Rij et al., 2020).

The GAMM analysis revealed that there was a significant main effect of the parametric term of place of articulation, *t*(1,535) = -50.0, *p* <.00001, indicating that there was a significant difference in the overall spectral densities of the alveolar [tsʰuə́n] and retroflex [ʈʂʰuén] frication of our Singapore Mandarin speaking participants. Three frequency windows of significant difference between the modelled alveolar [tsʰuə́n] and retroflex [ʈʂʰuén] spectra were found with the *plot_diff* function: 668.2 – 890.0Hz, 1336.4 – 3564Hz, and 4009.1 – 22050.0Hz. This indicates that there were significant differences in not only the overall spectral density, but also in the shape of the modelled trajectories of the alveolar [tsʰuə́n] and retroflex [ʈʂʰuén] frication spectra as produced by our Singapore Mandarin speaking participants.

6.2.5 Discussion

As in our earlier documentation on production of the Singapore Mandarin alveolar-retroflex contrast (in Chapter 4), the results of the speech production portion of this study once again shows that Singapore Mandarin speakers do produce alveolar and retroflex frication with a clear acoustic distinction. This is even found for frication in the minimal pair of words [tsʰuə́n] and [ʈʂʰuén] which have been documented to have higher levels of acoustic ambiguity in speech production due to coarticulatory effects from the low rounded vowel /uə/ (Chang, 2012; Jeng, 2006). At the same time, there are observable differences in our participants' speech productions, with some participants demonstrating higher degrees of articulatory contrast as compared to others. This can be seen from a visual inspection of Figure 6.5.

6.3 Speech Perception: Spring-Village CROWN Game

Following the production task, participants then took part in the Spring-Village CROWN Game, run on Open Sesame (Mathôt et al., 2012). This task was identical in methodology to the Spring-Village CROWN game of Chapter 5.

6.3.1 Methods

6.3.1.1 Participants

The participants of the speech perception task were identical to the participants of the speech production task.

6.3.1.2 Equipment

The Spring-Village CROWN Game was conducted on a 10-inch LCD screen on a Microsoft Surface Go computer. All stimuli were presented in OpenSesame (Mathôt et al., 2012). Audio stimuli were presented to the participants with a pair of Audio-Technica Over-ear Monitoring Headphones ATH-M40x.

6.3.1.2 Stimuli

Visual stimuli consisted of cartoon depictions of a monkey, and of the two target words in Mandarin Chinese: a village for [tsʰuə́n], and a spring festival (also known as Lunar New Year) for [ʈʂʰuə́n]. Figure 5.1 contains an example of the visual stimuli presented to participants.

Audio stimuli consisted of eight 580ms acoustic tokens synthesised along an 8-step Singapore Mandarin alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] (村 ‘village’ - 春 ‘spring’) continuum. Each token on the continuum was spaced equally from its adjacent tokens terms of acoustic distance. The alveolar end of the continuum (i.e., the most alveolar [tsʰuə́n]-like sound) was represented by Sound 1, and the retroflex end of the continuum (i.e., the most retroflex [ʈʂʰuə́n]-like sound) was represented by Sound 8.

6.3.2 Procedure

The procedure for the Spring-Village CROWN Game and the Language ID for the speech perception portion of this study were identical to that of the Spring-Village CROWN Game of Chapter 5.

Participants were first introduced to the task with written instructions in English. The instructions followed a storyline in which participants were told they needed to help a little monkey collect flowers for a flower crown by directing the little monkey travel to one of two locations (i.e., to the village or to the spring festival) based on the word that they heard on each trial.

6.3.2.1 Practice phase

To ensure that participants were familiar with the word sound to picture pairing, participants first completed two practice blocks. On each practice trial, participants were presented with the little monkey in the middle of the screen, and the pictures of the two locations (i.e., the village and the spring festival) on either side of the screen, while either Sound 1 or Sound 8 was played (see Figure 5.1 for example of visual stimuli). The participant would then make a keypress response to direct the monkey to either side of the screen. Following a keypress response, written feedback in English would be provided at the top of the screen (see Figure 5.2 for example of feedback provided following each practice response). Each sound token was played two times in random order over the course of each block. To maximise participants' likelihood of being able to remember the sound to picture pairing, the two locations (i.e., the village and the spring festival) were not counterbalanced across blocks.

6.3.2.2 Test phase

Following completion of the practice phase, participants completed a total of 10 test blocks. The test blocks were similar to the practice phase, with the exception that all eight sound tokens from the 8-step Singapore Mandarin alveolar-retroflex continuum were played in random order across each block, and no written feedback was given following a keypress response so as to not influence the participants' response patterns.

6.3.3 Analysis Plan

The analysis plan for the Spring-Village CROWN Game responses was identical to that of the analysis plan in Chapter 5. We obtained two measures of Mandarin alveolar-retroflex contrast perception from the response patterns of our participants in the Spring-Village CROWN game. First, we determined the average boundary threshold on the 8-step Mandarin alveolar-retroflex [tsʰuə́n]-

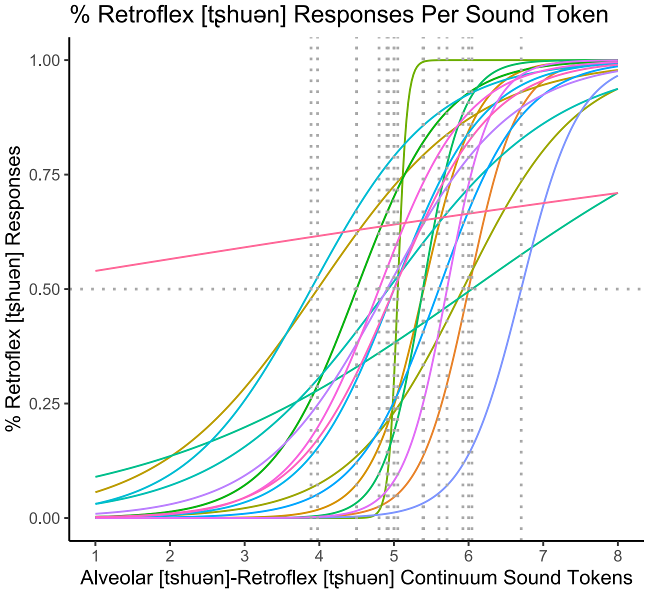
[ʈʂʰuə́n] continuum at which the majority of the participants' sound token identification decisions switched from the alveolar-initial word [tsʰuə́n] to the retroflex-initial word [ʈʂʰuə́n]. Next, we obtained the slope value of the decision gradients of each participants' response patterns for each sound token on the continuum. This decision gradient slope values represent the manner in which participants resolve perceptual ambiguity for speech tokens falling between the prototypical alveolar Sound 1 [tsʰuə́n] and retroflex Sound 8 [ʈʂʰuə́n]. Steep slopes represent a highly systematic response pattern consistent with a high degree of contrast sensitivity for perceptually ambiguous sound tokens the continuum, and shallow slopes represent a less systematic response pattern consistent with a lower degree of contrast sensitivity for perceptually ambiguous sound tokens on the continuum.

To obtain the boundary threshold and decision gradient slope values, we obtained the number of "spring festival" (Sound 8 retroflex [ʈʂʰuə́n]) responses each participant made for each of the eight sound tokens on the alveolar-retroflex contrast continuum across the test phase. Psychometric functions were then fit for each participant with *quickpsy* function in R (Linares & López-Moliner, 2016; R Core Team, 2020) based on the proportion of "spring festival" (Sound 8 retroflex [ʈʂʰuə́n]) responses made out of all the responses for each of the eight tokens on the alveolar-retroflex contrast continuum. Individual slope and boundary threshold values were then extracted for each participant from their fitted psychometric functions. Following this, we then obtained the median boundary threshold of all our participants. This median boundary threshold represents the median Mandarin alveolar-retroflex boundary threshold for our Singapore Mandarin speakers.

6.3.4 Results

Figure 6.8

Individual fitted decision gradient slopes of each participant.



Note. *N* = 20. Dotted lines indicate individual perceptual boundary thresholds.

The individual decision gradient slope values of each participant are shown in Figure 6.8 together with each participant's alveolar-retroflex boundary threshold (represented by grey dotted lines). From the fitted psychometric functions, we obtained an overall median boundary threshold of 5.0 (*SD* = 1.46).

6.3.5 Discussion

The results of this Spring-Village CROWN game revealed that the majority of our participants showed decision gradient responses characteristic of perception of a distinct Mandarin alveolar-retroflex phoneme contrast. This is similar to the results of our earlier Spring-Village CROWN game in Chapter 5. However, visual inspection of Figure 6.8 shows that our participants had a range of individual differences in their response patterns. In particular, the steepness of the alveolar-retroflex discrimination pattern differed between participants. One participant showed little evidence of systematic discrimination of the sound tokens (pink shallow slope), and one participant showed a highly systematic discrimination pattern for the sound tokens heard in the task (green steep slope).

The majority of participants' boundary thresholds were clustered around the midpoint of the continuum, and most participants' perceptual identification responses switched over from the alveolar to the retroflex at sound token 5. However, individual differences can still be seen, with some participants having a boundary threshold closer to the alveolar or retroflex end of the spectrum. This shows that there are individual differences in the degree of acoustic ambiguity that participants accept in a representation of an alveolar or retroflex sound. Participants with boundary thresholds closer to the alveolar end of the spectrum accepted a wider range of acoustic ambiguity as being consistent with retroflexion, while participants with boundary thresholds closer to the retroflex end of the spectrum accepted a wider range of acoustic ambiguity as being consistent with alveolar frication.

6.4 Perception-Production Links in Speech?

From the results of our participants' perception and production of the Singapore Mandarin alveolar-retroflex word pair [tsʰuə́n] and [ʈʂʰuə́n], we can see that there is a range of individual differences in how contrastively the alveolar and retroflex frication are perceived and produced. There are also observable differences in perceptual alveolar-retroflex boundary thresholds, and the densities of speech productions of alveolar and retroflex frication. We thus examined this data to determine if we could find any links within the individual differences of our participants' perception and production of frication in the Singapore Mandarin alveolar-retroflex word pair [tsʰuə́n] and [ʈʂʰuə́n].

6.4.1 Analysis Plan

We collapsed across gender for all analyses as our sample size of male participants was very small ($N = 4$). As the decision gradient slope values and the perceptual boundary thresholds obtained for each participant in this study were not normally distributed, a log transform was carried out on all decision gradient slope values and perceptual boundary thresholds to approximate Gaussian distribution. One female participant (Participant 22) was excluded from the perception-production analyses as the participant showed little evidence of being able to discriminate between the words [tsʰuə́n] and [ʈʂʰuə́n], and their extremely low boundary threshold data had an influential outlier effect on the boundary threshold analyses that we conducted.

6.4.1.1 Perception-production link analyses with CoG values

We conducted three linear mixed models analyses on the speech production CoG data to examine links between our participants’ perception and production of frication in the Singapore Mandarin alveolar-retroflex word pair [tsʰuə́n] and [ʈʂʰuə́n].

The first model assessed links between the degree of contrast in our participants’ perception and production of the alveolar-retroflex phoneme contrast. We used our participants’ Spring-Village CROWN Game decision gradient slope values as a measure of their phoneme contrast sensitivity. As a measure of speech production contrastiveness, CoG Δ values were obtained for each participant. CoG values were averaged over each place of articulation for each participant, and individual CoG Δ values were obtained from the difference between the averaged alveolar and retroflex CoG values.

A linear mixed effects model analysis was conducted in R (R Core Team, 2020) with the *lme* function from the *nlme* package (Pinheiro et al. 2022) on the CoG Δ values with a fixed factor of decision gradient slope value, and random intercept and slope of participant.

The second model assessed links between our participants’ perceptual alveolar-retroflex boundary thresholds and the CoG values of each of their alveolar frication productions. A linear mixed effects model analysis was conducted in R (R Core Team, 2020) with the *lme* function from the *nlme* package (Pinheiro et al., 2022) on all participants’ alveolar [tsʰuə́n] frication CoG values with a fixed factor of perceptual boundary threshold, and random intercept of participant random by-participant slope for trial number.

The third model assessed links between our participants’ perceptual alveolar-retroflex boundary thresholds and the CoG values of each of their retroflex frication productions. A linear mixed effects model analysis was conducted in R (R Core Team, 2020) with the *lme* function from the *nlme* package (Pinheiro et al. 2022) on all participants’ retroflex [ʈʂʰuə́n] frication CoG values with a fixed factor of perceptual boundary threshold, and random intercept of participant and random by-participant slope for trial number.

6.4.1.2 Perception-production link analyses with LTAS GAMMs

We then conducted three general additive mixed model (GAMM) analyses to examine links between our participants’ perception and production of frication in the Singapore Mandarin alveolar-retroflex word pair [tsʰuə́n] and [ʈʂʰuə́n].

As with the CoG analyses, the first model assessed links between the degree of contrast in our participants’ perception and production of the alveolar-retroflex phoneme contrast. We used our participants’ Spring-Village CROWN Game decision gradient slope values as a measure of their phoneme contrast sensitivity. As a measure of speech production contrastiveness, we obtained alveolar-retroflex difference matrices for each participant. The LTAS matrices of each frication token were interpolated across 100 frequency bins from 0-22050Hz, and spectral data were averaged across each place of articulation for each participant. An alveolar-retroflex difference matrix was obtained for each participant from the difference between their averaged alveolar and retroflex LTAS matrices. Using the *bam* function from the *mgcv* package (Wood, 2011) in R (R Core Team, 2020), we then conducted a GAMM analysis on the alveolar-retroflex differences matrices with the parametric fixed term of place of decision gradient slope value, smooth term of decision gradient slope value and random smooths for participant and trial number.

The second model assessed links between our participants' perceptual alveolar-retroflex boundary thresholds and the LTAS matrices obtained for each of their alveolar [tsʰuə́n] frication productions. Using the *bam* function from the *mgcv* package (Wood, 2011) in R (R Core Team, 2020), we conducted a GAMM analysis on our participants' alveolar [tsʰuə́n] frication LTAS data with the parametric fixed term of place of perceptual boundary threshold, smooth terms of perceptual boundary threshold and random smooths for participant and trial number.

The third model assessed links between our participants' perceptual alveolar-retroflex boundary thresholds and the LTAS matrices obtained for each of their retroflex[ʈʂʰuə́n] frication productions. Using the *bam* function from the *mgcv* package (Wood, 2011) in R (R Core Team, 2020), we conducted a GAMM analysis on our participants' retroflex [ʈʂʰuə́n] frication LTAS data with the parametric fixed term of place of perceptual boundary threshold, smooth term of perceptual boundary threshold, and random smooths for participant and trial number.

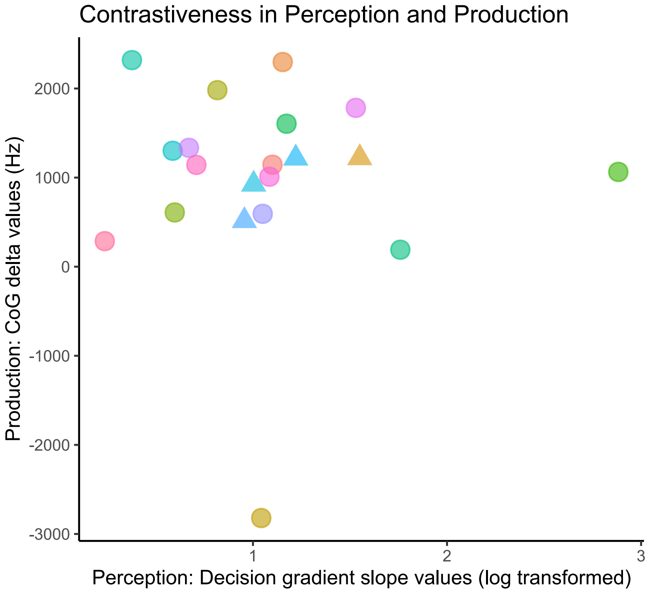
Based on Sóskuthy (2017), a significant effect of a parametric term indicates a significant difference in the overall spectral density of the modelled data for each condition of the parametric term. Thus, we reported the effects of the parametric terms assessed in each of the models. To ensure that the basis dimensions of the GAMMs would not result in oversmoothing of the fitted data, we conducted model criticism checks for each GAMM analysis using the *gam.check* function of the *mgcv* package (Wood, 2011). Based on the criticism checks, we selected k-values 70 for the first GAMM analysis, and 30 for the second and third GAMM analyses.

6.4.2 Results

6.4.2.1 Perception-production results with CoG values

Figure 6.9

Individual [tsʰuə́n]-[ʈʂʰuə́n] CoG Δ values plotted against decision gradient slope values.

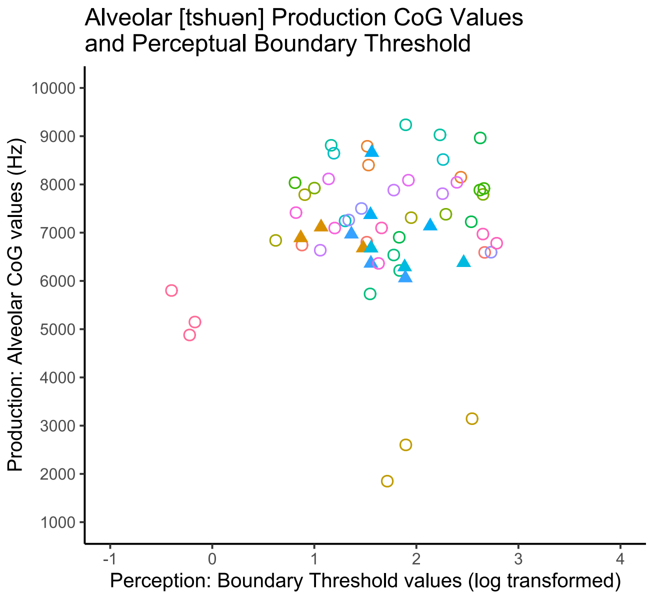


Note. *N* = 20. Circles: female; triangles: male. Slope values jittered by .25 for visualisation. Participant 22 – pink circle – excluded from statistical analysis.

The first linear mixed effects model analysis conducted on the CoG Δ values revealed that there was no significant main effect of decision gradient slope values, $t(1,17) = -.31$, $p = .76$, $\eta p^2 < .0001$, indicating that there was no significant link between constrastiveness of perception and production of frication in the alveolar-retroflex affricate contrast in [tsʰuə́n] and [ʈʂʰuén]. Figure 6.9 shows participants' perceptual decision gradient slope values plotted against their CoG Δ values.

Figure 6.10

Individual alveolar [tsʰuə́n] frication CoG values plotted against perceptual alveolar-retroflex boundary thresholds.

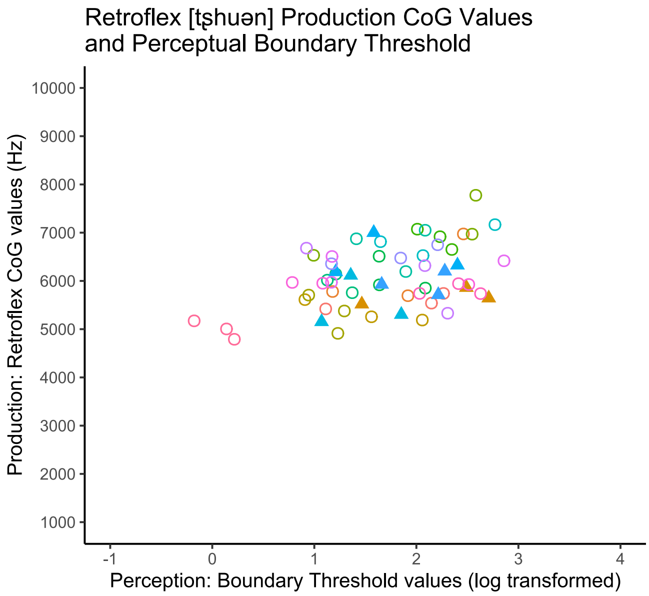


Note. *N* = 20. Circles: female; triangles: male. Boundary threshold values jittered by .25 for visualisation. Participant 22 – pink circles – excluded from statistical analysis.

The second linear mixed effects model analysis conducted on the alveolar [tsʰuə́n] frication CoG values revealed that there was no significant main effect of perceptual boundary threshold, $t(1,17) = 2.10$, $p = .051$, $\eta p^2 = .21$, indicating that there was no significant link between the degree of acoustic ambiguity participants accepted for representations of alveolar frication, and their production of alveolar frication as measured by CoG values. Figure 6.10 shows participants’ perceptual boundary threshold values plotted against their alveolar CoG values.

Figure 6.11

Individual retroflex [t͜ʂʰuə́n] frication CoG values plotted against perceptual alveolar-retroflex boundary thresholds.



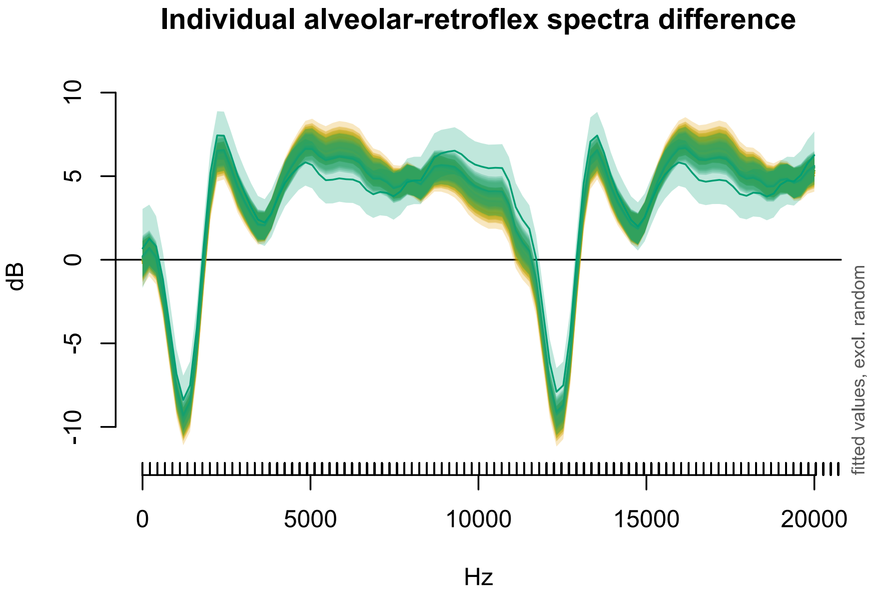
Note. *N* = 20. Circles: female; triangles: male. Boundary thrshold values jittered by .25 for visualisation. Participant 22 – pink circles – excluded from statistical analysis.

However, the third linear mixed effects model analysis conducted on the retroflex [t͜ʂʰuə́n] frication CoG values revealed that there was a significant main effect of perceptual boundary threshold, *t*(1,17) = 3.16, *p* = .0057, ηp^2 = .37 indicating that there was a significant link between the degree of acoustic ambiguity participants accepted for representations of retroflex frication, and their production of retroflex frication as measured by CoG values. Figure 6.11 shows participants' perceptual boundary threshold values plotted against their retroflex CoG values.

6.4.2.2 Perception-production link results with LTAS GAMMs

Figure 6.12

Models of alveolar [tsʰuə́n]-retroflex [ʈʂʰuén] frication spectra difference matrices fitted individually for each participant.

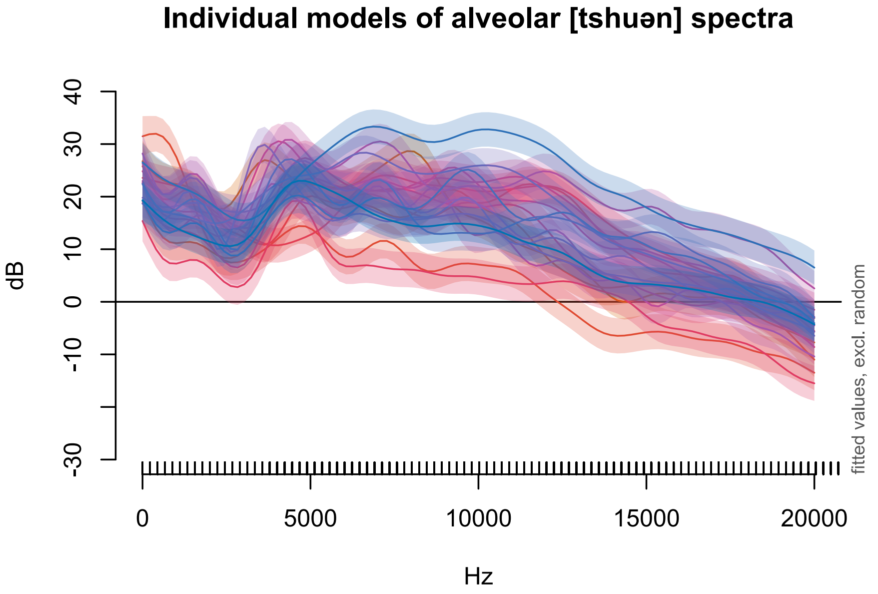


Note. *N* = 19. Lines coloured by decision gradient slope value on an orange-green gradient. Orange: lower decision gradient slope values. Green: higher decision gradient slope values.

The first GAMM analysis conducted on the alveolar-retroflex differences matrices revealed that there was no significant main effect of the parametric term of decision gradient slope value, *t*(1,64) = -0.26, *p* = .80, indicating that there was no significant link between constrastiveness of perception and overall spectral density differences in the production of frication in the alveolar-retroflex affricate contrast in [tsʰuə́n] and [ʈʂʰuén]. Figure 6.12 shows the modelled difference matrices of each participant, with each trajectory coloured by decision gradient slope value on an orange to green gradient (orange: lower decision gradient slope values, green: higher decision gradient slope values).

Figure 6.13

Models of alveolar [tsʰuə́n] frication spectra fitted individually for each participant.

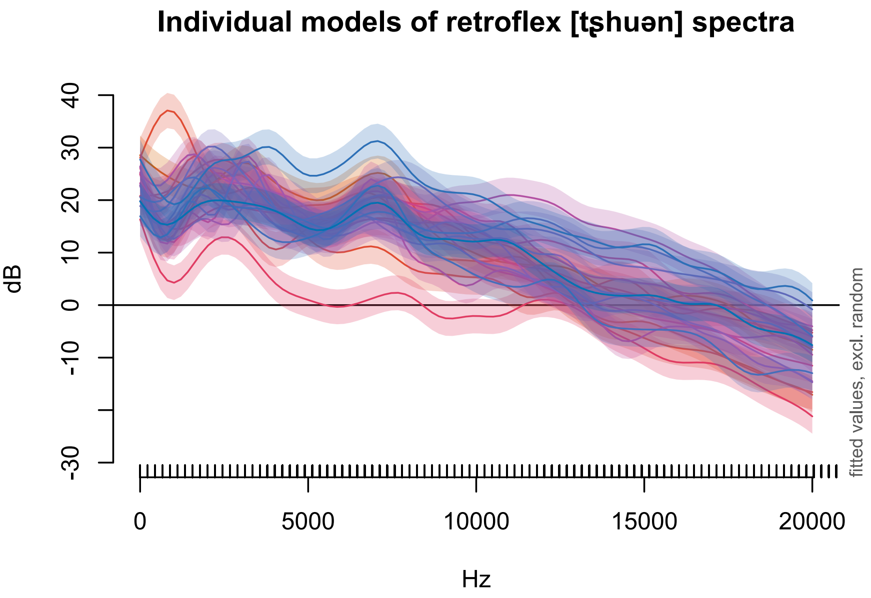


Note. *N* = 19. Lines coloured by boundary threshold on a red-blue gradient. Red: boundary thresholds closer to the alveolar end of the continuum, Blue: boundary thresholds closer to the retroflex end of the continuum.

The second GAMM analysis conducted on the alveolar [tsʰuə́n] frication LTAS data revealed that there was no significant main effect of the parametric term of perceptual boundary threshold, *t*(1,1200) = .40, *p* = .69, indicating that there was no significant link between the degree of acoustic ambiguity participants accepted for representations of alveolar frication, and the overall spectral density of their productions of alveolar frication. Figure 6.13 shows the modelled alveolar frication spectra of each participant, with each trajectory coloured by boundary threshold on a red to blue gradient (red: boundary thresholds closer to the alveolar end of the continuum, blue: boundary thresholds closer to the retroflex end of the continuum).

Figure 6.14

Models of retroflex [ʈʂʰuén] frication spectra fitted individually for each participant.



Note. *N* = 19. Lines coloured by boundary threshold on a red-blue gradient. Red: boundary thresholds closer to the alveolar end of the continuum, Blue: boundary thresholds closer to the retroflex end of the continuum.

However, the third GAMM analysis conducted on the retroflex [ʈʂʰuə́n] frication LTAS data revealed that there was a significant main effect of the parametric term of perceptual boundary threshold, *t*(1,1192) = 2.72, *p* = .0065, indicating that there was a significant link between the degree of acoustic ambiguity participants accepted for representations of retroflex frication, and the overall spectral density of their productions of retroflex frication. Figure 6.14 shows the modelled retroflex frication spectra of each participant, with each trajectory coloured by boundary threshold on a red to blue gradient (red: boundary thresholds closer to the alveolar end of the continuum, blue: boundary thresholds closer to the retroflex end of the continuum).

6.4.3 Discussion

Our investigation of perception-production links in speech for a Singapore Mandarin alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast showed that there does not appear to be a link between the degree of contrast sensitivity that our Singapore Mandarin speakers have for the Mandarin alveolar-retroflex contrast, and the degree of contrastiveness that is present in acoustic measurements of their production of alveolar and retroflex frication. This appears to indicate that high levels of phoneme contrastiveness in speech production are not required for Singapore Mandarin speakers to demonstrate high levels of structured perceptual sensitivity for the Mandarin alveolar-retroflex contrast. This pattern was consistent between analyses using both CoG values as well as LTAS data as measurements of alveolar-retroflex frication in speech production as no significant links were found for either analysis

At the same time, our analyses also showed that there does not appear to be a link between the degree of perceptual ambiguity that participants judged to be consistent with their representation of a voiceless alveolar affricate in the Singapore Mandarin [tsʰuə́n], and individual differences in their speech productions of the same voiceless alveolar affricate. This pattern was consistent between analyses using both CoG values as well as LTAS data as measurements of alveolar frication in speech production as no significant links were found for either analysis.

However, we did find a link between the degree of perceptual ambiguity that participants judge to be consistent with their representation of a voiceless retroflex affricate in the Singapore Mandarin [ʈʂʰuə́n], and individual differences in their speech productions of the same voiceless retroflex affricate. Looking at the perception-production analysis on the retroflex [ʈʂʰuə́n] frication CoG values (see Figure 6.11), we can see that there appears to be a trend in which participants who articulate their retroflex frication more closely to the alveolar in acoustic space (i.e., higher CoG values) are also more likely to judge ambiguous sounds as alveolar rather than retroflex (i.e., with perceptual boundary thresholds closer to the retroflex end of the spectrum). At the same time, a similar trend can also be seen in the perception-production GAMM analysis conducted on the retroflex [ʈʂʰuə́n] frication LTAS data. A visual inspection of Figure 6.14 shows that participants with perceptual boundary thresholds closer to the retroflex end of the spectrum appeared to also have spectral trajectories for retroflexion that were closer in acoustic space to the alveolar trajectories shown in Figure 6.13. As such, both the CoG and GAMM analyses revealed that there was a significant link between perceptual boundary thresholds and retroflexion in speech production, with higher boundary thresholds being linked to the production of retroflex frication that is more alveolar-like.

This pattern of results appears to show that participants who display more deretroflexion in speech tend to have lower levels of sensitivity for retroflex cues in ambiguous speech tokens. Instead, they appear to only identify less ambiguous stimuli as acceptable examples of retroflexion. This aligns with studies on second language acquisition and specific articulatory disorders which demonstrate that individuals typically have lower levels of perceptual sensitivity for phonemes that they do not produce as clearly (Bradlow et al., 1997; Rvachew & Jamieson, 1989; Sakai & Moorman, 2018). With a larger, more representative sample size, we will be able to determine if this link is generalisable to the wider population of Singapore Mandarin speakers. At the same time, studies have found that perceptual training on phoneme contrasts can lead to concomitant changes in production of the phoneme contrasts (e.g., Bradlow et al., 1997; Inceoglu, 2016; Sakai & Moorman, 2018). It would be of interest to determine if perceptual training on this Singapore Mandarin alveolar-retroflex contrast might lead to changes in speech production.

On the other hand, the link that we found between retroflexion in speech and perceptual boundary thresholds appears to contrast with findings in existing studies on Mandarin speakers. Chang et al. (2013) suggested that deretroflexion in Taiwan Mandarin speakers is linked to a higher degree of tolerance for perceptual ambiguity in retroflexion – this is the opposite of what we found with our participants. However, this difference in our findings could be related to the fact that Chang et al. (2013) assessed perceptual sensitivity in their Taiwan Mandarin speaking participants with ambiguous alveolar-retroflex speech tokens synthesised based on standard Beijing Mandarin. In an earlier study, Chang (2012) found that the distance in acoustic space for the Mandarin alveolar-retroflex contrast is typically larger in Beijing Mandarin speakers as compared to Taiwan Mandarin speakers. Chang (2012) also found that Taiwan Mandarin speakers tend to have higher retroflex CoG values as compared to Beijing Mandarin speakers. Thus, it is possible that the Taiwan Mandarin speakers tested by Chang et al. (2013) may have accepted a larger degree of ambiguity in *Beijing* Mandarin as being consistent with their mental representations of retroflexion, as the acoustic characteristics of the ambiguous speech tokens may have in fact been more closely aligned to the acoustic norms of retroflex frication rather than alveolar frication in their native variety of *Taiwan* Mandarin. This demonstrates that speech perception tests should take into consideration the unique characteristics of the native language variety of its participants.

6.5 Conclusion

Overall, this study revealed that there does indeed appear to be a link between speech perception and production of a Singapore Mandarin alveolar-retroflex affricate [tsʰuə́n]-[ʈʂʰuə́n] contrast on the level of the individual. In particular, we discovered that participants whose retroflex speech productions were closer to alveolar frication in acoustic space tend to show less sensitivity for retroflex cues in ambiguous stimuli, and required higher levels of acoustic clarity in order to perceive sound tokens on a Singapore Mandarin [tsʰuə́n]-[ʈʂʰuə́n] continuum as acceptable examples of Mandarin retroflexion. We did not find significant evidence for any similar link in our participants' alveolar speech productions, suggesting that the perception-production link of this [tsʰuə́n]-[ʈʂʰuə́n] contrast in Singapore Mandarin may be more closely tied to the retroflex place of articulation rather than the alveolar. At the same time, we did not find any links between degrees of contrastiveness in perception and production of the Singapore Mandarin [tsʰuə́n]-[ʈʂʰuə́n] contrast, indicating that general distance in acoustic space of Mandarin alveolar-retroflex speech productions does not appear to be related to how the level of perceptual sensitivity that participants have for a Singapore Mandarin alveolar-retroflex contrast.

Further investigation can be conducted with a larger, more representative sample size to determine if this perception-production link we found can be generalised to the larger population of Singapore Mandarin speakers. We also plan to examine the effect of perceptual sensitivity training on this Singapore Mandarin [tsʰuə́n]-[ʈʂʰuə́n] contrast to determine if changes in fine-grained perceptual sensitivity might carry over to changes in speech production. This would be of value in education settings, as well as in speech therapy settings to assist individuals in learning to perceive and produce speech sounds. A suitable perceptual training paradigm would be of great value for this future investigation. We have developed such a paradigm based on distributional learning which will be discussed in our next chapter.

Over the course of the studies discussed so far, we have created a well-validated stimulus set that is aligned to the acoustic norms of the Singapore Mandarin [tsʰuə́n]-[ʈʂʰuə́n] alveolar-retroflex affricate contrast. At the same time, we have also developed a well-validated means of assessing fine-grained perceptual sensitivity for this contrast with the Spring-Village CROWN Game. With this set of stimuli and the Spring-Village CROWN Game, we plan to effectively investigate the hypothesis of a language-specific transfer leaning effect in Singapore English-Mandarin bilingual adults.

Chapter 7: Investigating a language-specific learning advantage with a synthesised alveolar-retroflex sibilant contrast based on Singapore Mandarin (distributional learning)

7.1 Introduction

In the very first study reported for this project, we attempted to address the question of whether language learning advantages in bilinguals might be linked to language-specific advantages based on transfer effects from overlaps between known and to be learnt language stimuli. For the purposes of our first study on this question, we chose to utilise a bimodal distributional training paradigm to train Singapore English-Mandarin bilingual adults on a novel 7-step Hindi dental-retroflex /d̪a/-/ɖa/ contrast continuum taken from a series of learning studies conducted with children and adults (Golestani & Zatorre, 2004; Golestani & Zatorre, 2009; Vandermosten et al. 2018). We based our training stimulus decision on an understanding that higher levels of Mandarin understanding abilities should be linked to language-specific transfereffects for the novel Hindi dental-retroflex contrast as Hindi and Mandarin shared a dental/alveolar-retroflex place of articulation contrast, albeit with a different manner of articulation (stops for Hindi, sibilants for Mandarin). However, we were unable to observe any evidence for a language-specific transfer effect in our first study, as we did not obtain an overall distributional learning effect in a large, representative sample of our target population.

We reasoned that the results we obtained in our first study were likely due to two main factors. Firstly, the synthesised Hindi dental-retroflex continuum created by Golestani and Zatorre (2004) was likely to be too difficult for adults to learn via distributional learning. In fact, a native Hindi speaker who participated in our first study reported being unable to tell the /d̪a/ and /ɖa/ continuum endpoint tokens apart, and also showed no distributional learning effect following bimodal distributional training. While studies using this set of stimuli have managed to find some learning effects in adults, these studies utilise a much longer “perceptual fading” training paradigm lasting around an hour with trial-by-trial feedback – this paradigm may have been more suitable for adults who are typically poorer at extracting distributional cues from non-native phoneme stimuli (e.g., Wanrooij et al. 2014).

Secondly, the Singapore variety of Mandarin does not, in fact, utilise the same place of articulation as in Hindi for the production of phonemes typically described as a “retroflex”, and our lingual ultrasound study examining Singapore Mandarin retroflexion showed that our sample of Singapore Mandarin speakers typically produce Mandarin “retroflex” phonemes with a “humped” laminal alveopalatal place of articulation, unlike the “curled” retroflex found in Hindi. The lack of structural overlap between the feature of retroflexion in Hindi and Singapore Mandarin, may explain why we were unable to observe evidence of any language-specific transfer effects amongst Singapore English-Mandarin bilingual adults.

In order to accurately assess such language-specific learning effects in our target population, we needed to ensure that the training stimuli we used would actually be aligned to the acoustic norms of the local population of Singapore Mandarin speakers. Such a stimulus set did not yet exist at the start of our project. While there are existing large-scale corpora of Mandarin words such as The Chinese Lexicon Project (Sze et al., 2014) and the two-character compound word Chinese Lexicon Project (Tse et al., 2017), these corpora are compilations of the typical speech of speakers of standard Mainland Mandarin (Sze et al., 2014), and Hong Kong Mandarin speakers (Tse et al., 2017), and would not be well-suited for Singapore Mandarin speakers. We know from existing studies that the Singapore variety of Mandarin is distinct from the standard variety Mandarin, particularly in the kind of retroflexion in speech due to influence from non-retroflex varieties of Chinese languages such as Hokkien (e.g., Chen et al., 2016; Lock, 1989; Ng, 1985). Thus, we set out to synthesise our own set of stimuli suitable for use in a distributional learning study of Singapore Mandarin speakers. This would allow us to more accurately assess language-specific transfer effects in our target population.

We thus created our own set of stimuli consisting of an 8-step synthesised alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast continuum based on the acoustic norms of 42 Singapore Mandarin speaking adults (Chapter 4). We also developed a study paradigm (the Spring-Village CROWN Game) that we have validated in 62 Singapore Mandarin speaking participants (Chapter 5) to be sensitive enough to capture fine-grained individual differences in perceptual contrast sensitivity for our synthesised alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast continuum. As such, we were confident that we would be able to use this set of stimuli with a distributional learning adaptation of the Spring-Village CROWN Game to investigate if bilingual learning advantages are linked to language-specific transfer effects. In order to ensure that our stimuli would be suitable for eliciting distributional learning effects in Singapore English-Mandarin bilingual adults, we first conducted a small pilot study of *N* = 20 participants to determine if we could find overall distributional learning effects in the form of increased sensitivity for ambiguity embedded in our new synthesised alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast continuum with the use of a bimodal distributional training paradigm. Following this, we then conducted a preregistered (https://osf.io/ab8gp) large-scale study of *N* = 50 participants to determine if a distributional learning effect could be generalised to a larger, more representation sample size, and to assess for language-specific transfer effects. We hypothesised that we would find distributional learning effects for our participants, and we also hypothesised that if language-specific transfer effects exist, participants with stronger Mandarin abilities would have larger distributional learning effects for the synthesised alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast following bimodal distributional training.

To promote open science practices, we have made all relevant data, task files, stimuli, and analysis files for this study available on the Open Science Framework (https://osf.io/h9myq/).

7.2 Pilot Study

Before conducting a full-scale study to examine our two hypotheses, we first conducted a pilot study of *N* = 20 to ensure that bimodal distributional training on our synthesised Singapore Mandarin alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast would indeed elicit overall distributional learning effects bilingual adults of the Singapore varieties of English and Mandarin. As in Chapter 2, the term ‘distributional learning effects’ will be used to refer to changes in perception of a phoneme contrast following training on a bimodally distributed frequency of sounds drawn from a continuum based on the target contrast. In this chapter, the term ‘distributional learning effects’ will specifically refer to increases in perceptual sensitivity for ambiguity embedded in a synthesised Singapore Mandarin alveolar-retroflex contrast continuum following bimodal distributional training on sounds drawn from the continuum.

7.2.1 Methods

7.2.1.1 Participants

All participants who took part in our perception-production study in Chapter 6 were invited to take part in the distributional learning study. All participants except for one male participant agreed to participate in this pilot study. As such, 20 participants (16 female, 4 male) took part in this study in exchange for $5 for every half hour of their participation. Participants were aged between 20 – 33 years of age (Median = 23.5, *SD* = 3.8). All participants reported being English-Mandarin bilinguals, and having grown up mostly in Singapore. The study was reviewed by the ethics board of the university (NTU-

IRB-2019-01-034) and all participants gave informed consent prior to taking part in the study tasks.

7.2.1.2 Equipment

This pilot study was conducted on a 10-inch LCD screen on a Microsoft Surface Go computer. All stimuli were presented in OpenSesame (Mathôt et al., 2012). Audio stimuli were presented to the participants with a pair of Audio-Technica Over-ear Monitoring Headphones ATH-M40x.

7.2.1.3 Stimuli

The visual stimuli used in the practice and test phases of this pilot study were identical to that of the Spring-Village CROWN game tasks conducted in our earlier studies. As such, visual stimuli consisted of cartoon depictions of a monkey, and of the two target words in Mandarin Chinese: a village for [tsʰuə́n], and a spring festival (also known as Lunar New Year) for [ʈʂʰuə́n]. Figure 5.1 contains an example of the visual stimuli presented to participants. The visual stimuli used during the unattended bimodal distributional training phase of the study consisted of muted cartoon clips (selected clips from Series 2 of Mr Bean: The Animated Series; Atkinson et al., 2003).

The audio stimuli used in this task were identical to those of the Spring-Village CROWN Game conducted in our earlier studies (Chapters 5 and 6). As such, audio stimuli consisted of eight 580ms acoustic tokens synthesised along an 8-step Singapore Mandarin alveolar-retroflex [tsʰuə́n] - [ʈʂʰuə́n] (村 ‘village’

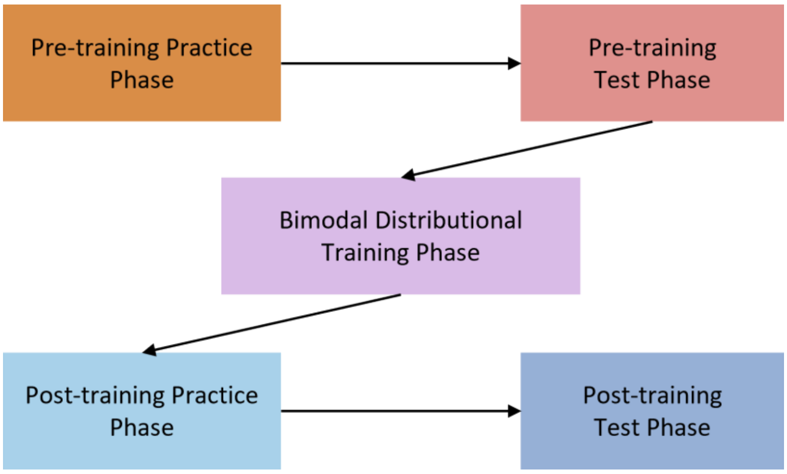
- 春 ‘spring’) continuum. Each token on the continuum was spaced equally from its adjacent tokens in terms of acoustic distance. The alveolar end of the continuum (i.e., the most alveolar [tsʰuə́n]-like sound) was represented by Sound 1, and the retroflex end of the continuum (i.e., the most retroflex [ʈʂʰuə́n]-like sound) was represented by Sound 8. During the unattended bimodal distributional training phase of the study, participants were presented with a series of sound tokens drawn from the 8-step Singapore Mandarin alveolar-retroflex contrast continuum.

7.2.2 Procedure

The procedural flow of the study is shown in Figure 7.1.

Figure 7.1

Procedural flow of bimodal distributional learning study.



7.2.2.1 Pre- and post-training practice phase

To ensure that participants were familiar with the sound to picture pairing [tsʰuə́n]: ‘village’, [ʈʂʰuə́n]: ‘spring’ festival, participants completed two practice blocks prior to each test phase. The practice phases in this study were identical to those of the Spring-Village CROWN Game conducted in our earlier study (Chapter 6).

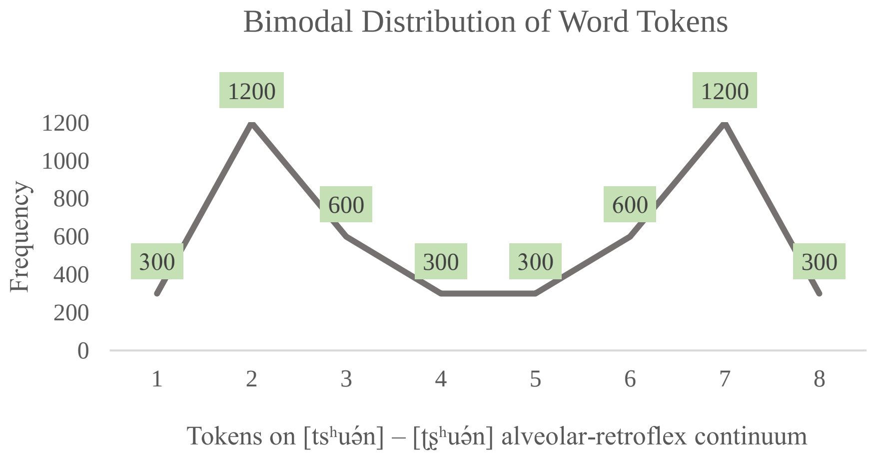
7.2.2.2 Pre- and post-training test phase

Participants took part in two test phases, one before the training phase, and one after the training phase. The training phases in this study were identical to those of the Spring-Village CROWN Game conducted in our earlier study (Chapter 6).

7.2.2.3 Bimodal distributional training phase

Figure 7.2

Bimodal distributional frequency of training sound tokens presented to participant.



All participants took part in one unattended bimodal distributional training phase. During the training phase, participants were instructed to watch muted cartoon clips while audio stimuli played over a pair of headphones. For this unattended training paradigm, participants heard a total of 4,800 sound tokens drawn from the 8-step Singapore Mandarin alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] continuum presented in a bimodally distributed frequency of sound tokens (see Figure 7.2 for distribution of sound tokens). Sound tokens were presented to the participants in a random order with an 80ms ISI between each sound. The training phase lasted approximated 30 minutes in total.

7.2.2.4 Language ID

To collect data on participants’ language understanding abilities, they were asked to rate how well they understood the four main languages in Singapore (English, Mandarin, Malay, and Tamil) on a scale of 0 to 100 using an onscreen sliding scale. Participants were also given an option to add in more languages or dialects to rate.

7.2.3 Analysis Plan

To obtain individual measures of contrast sensitivity for each participant, we obtained the slope values of the decision gradients of each participants’ response patterns for each sound token on the continuum at pre-training test phase, and post-training test phase. The decision gradient slope values represent the manner in which participants resolve perceptual ambiguity for speech tokens falling between the prototypical alveolar Sound 1 [tsʰuə́n] and retroflex Sound 8 [ʈʂʰuə́n]. Steep slopes represent a highly systematic response pattern consistent with a high degree of contrast sensitivity for perceptually ambiguous sound tokens the continuum, and shallow slopes represent a less systematic response pattern consistent with a lower degree of contrast sensitivity for perceptually ambiguous sound tokens on the continuum.

To do this, we obtained the number of alveolar [tsʰuə́n] (village) and retroflex [ʈʂʰuə́n] (spring) responses each participant made for each of the eight tokens on the contrast continuum, and then obtained the proportion of retroflex [ʈʂʰuə́n] responses made for each of the eight tokens on the continuum. We fit psychometric functions to the response patterns of each participant at pre-training test phase and post-training test phase, and extracted the slope values of each psychometric function using the *quickpsy* function in R (Linares & López-Moliner, 2016; R Core Team, 2020). Decision gradient slope values were computed separately for the pre- and post-training test blocks.

7.2.3.1 Pre-training control analysis

To ensure that exposure to the sound tokens presented prior to the bimodal distributional training phase had not led to any distributional learning effects, we first conducted a control analysis on the slope values of the first half and second half of the pre-training blocks to determine if significant changes in slope value had occurred prior to the training phase. Individual slope values were obtained separately for each participant for the first half and second half of the pre-training test blocks for comparison. As the sample size was small, a non-parametric Wilcoxon signed rank test was conducted to compare changes in slope values between the first half and second half of the pre-training blocks.

7.2.3.1 Distributional learning analysis

We then conducted our main analysis and assessed for distributional learning effects by comparing our participants' pre-training baseline slope values to their post-training slope values. This is aligned with the methodology of existing bimodal distributional learning studies (e.g., Chládková et al., 2022). As the pilot sample size was small, a non-parametric Wilcoxon signed-rank test was conducted to determine if there were significant changes in slope values between the pre- and post-training test blocks using the *wilcox.test* function from the *stats* package (R Core Team, 2020). We expected to find a significant increase in slope values at post-training test phase as compared to pre-training test phase, indicating the presence of an overall distributional learning effect.

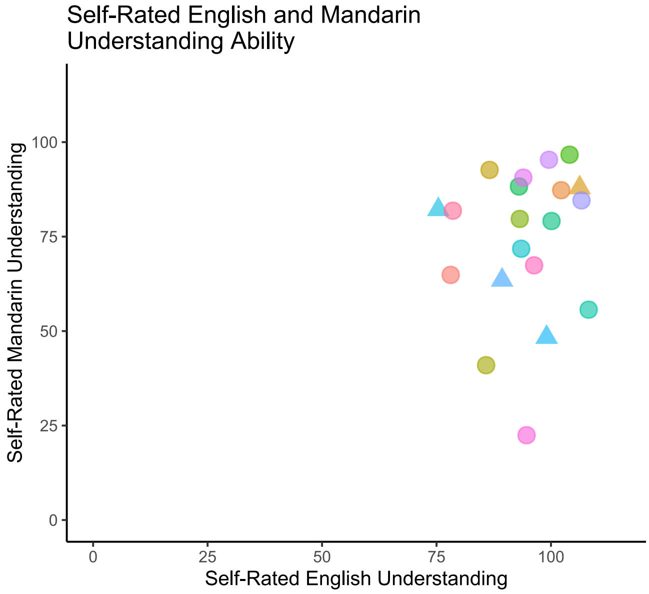
All analyses were carried out with the statistical software R (R Core Team, 2020).

7.2.4 Results

7.2.4.1 Language ID

Figure 7.3

Participants' self-reported English and Mandarin understanding ability.



Note. *N* = 20. Responses jittered by 10 for visualisation.

Overall, our pilot participants self-reported high levels of understanding abilities for spoken English (Range: 70 - 100, *Median* = 100, *SD* = 9.2), indicating that they would not have had any issues understanding the task instructions. Participants self-reported more varying levels of understanding abilities for spoken Mandarin (Range: 20 - 100, *Median* = 81, *SD* = 19.0). While one of our participants reported a Mandarin understanding ability that was much lower than our other participants (score of 20), all participants had taken part in formal education in Singapore and would have had received at least 10 years of formal education in Mandarin. Moreover, the words that we chose for our Mandarin speech production and perception tasks were easily understandable by young children. Therefore, we did not exclude this participant from our analyses. Figure 7.3 shows the distribution of our participants' self-rated English and Mandarin understanding abilities.

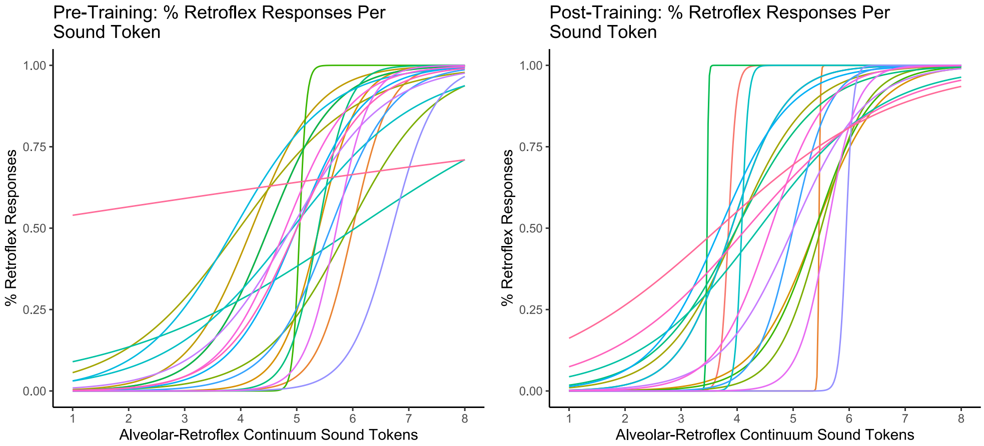
7.2.4.2 Pre-training control results

The control analysis revealed that there was no significant difference in decision gradient slope value between the first and second half of the pre-training phase, (*Z* = .64, *p* = .55, *r* = .14; pre-1: slope *M* = .70, *SD* = .33; pre-2: slope *M* = .68, *SD* = .25). This indicates that no learning had occurred prior to the bimodal distributional training phase.

7.2.4.2 Distributional learning results

Figure 7.4

Individual decision gradient slopes showing fitted psychometric functions for responses at pre-training test and post-training test.



Note. *N* = 20. Pre-training test: left. Post-training test: right.

Fitted psychometric slopes for participants' pre- and post-training responses can be seen in Figure 7.4, where a number of participants show visibly steeper slopes at the post-training as compared to the pre-training test block. The Wilcoxon signed-rank test revealed a significant difference in decision gradient slope values between pre-training and post-training test blocks, (*Z* = -1.98, *p* = .048, *r* = -.44; pre-training slope values: *M* = 2.5, *SD* = 3.4; post-training slope values: *M* = 12.1, *SD* = 24.9). This indicates that learning had occurred following the unattended bimodal distributional training phase.

7.2.5 Discussion

From the results of the pilot study, we can see that our participants did indeed have a significant overall distributional learning effect in the form of an overall increase in perceptual sensitivity following unattended bimodal distributional training on a synthesised Singapore Mandarin alveolar-retroflex [tsʰuə́n] - [ʈʂʰuə́n] continuum. Moreover, we can see clear distributional learning effects even for participants whose responses at pre-training test phase showed little sensitivity for contrastive perception of the different sound tokens (i.e., shallow slopes such as the pink line and green line in the pre-training section of Figure 7.4). As the distributional learning effect we obtained in the pilot study appears to extend to participants of different baseline ranges of contrast sensitivity, we were confident that this distributional learning effect would be generalisable to a larger, more representative sample size of Singapore English-Mandarin bilingual adult participants.

We thus conducted our next full-scale study with reason to believe that this study paradigm would allow us to test our experimental hypotheses. We expected that firstly, overall distributional learning effects can be found in Singapore English-Mandarin bilingual adults using our synthesised 8-step contrast continuum capturing sounds from the local variety of Singapore Mandarin. Secondly, we expected that we would be able to find evidence of a language-specific transfer effect amongst our participants, with stronger Mandarin understanding abilities being linked to larger distributional learning effects.

7.3 Main Study

Following the successful learning in the pilot study, we conducted our main full-scale study to investigate the hypothesis of language-specific transfer effects. As discussed in our first distributional learning study, familiarity with the feature of Mandarin retroflexion is likely to be associated with higher levels of Mandarin understanding abilities, as retroflexion is common in the standard variety of Mandarin typically used in formal education in Singapore (Chong & Tan, 2013). Therefore, if learning advantages in bilinguals stem from transfer effects from shared features between known languages and novel language stimuli (with more ambiguity than would be encountered in daily life), we reasoned that larger distributional learning effects should be found in individuals with higher levels of Mandarin understanding abilities. As in our earlier studies, we utilised our self-rated Language ID to obtain estimates of our participants' Mandarin understanding abilities. Studies have shown that self-assessment of language abilities can provide an accurate measure of linguistic ability in the absence of suitable empirical language proficiency tests for local varieties of languages. This is particularly so for self-rating language assessments conducted within a homogenous group of bilinguals, with self-rated proficiencies typically being correlated with empirical language proficiency tests (e.g., Thompson, 2015).

To ensure that our main study would be well-powered, we conducted an a priori power analysis in G*Power 3.1.9.7 to determine an appropriate minimum sample size, and established that 48 participants would be needed to observe an effect size equal to that of the pilot study (dz = .48) at an alpha level of .05 with a power of $1-\beta = 0.95$. We chose to increase our minimum sample size to $N = 50$ to ensure that we would have sufficient power for an additional supplementary exploratory factor analysis of language factors (de Winter et al., 2009). Additionally, we preregistered a data collection stop rule based on a Bayesian significance test of the distributional learning effects obtained. For this Bayesian test, we used a prior of .057 based on the log-transformed distributional learning effect obtained in the pilot study to determine if the distributional learning effect was substantially supportive of either H1 (BF > 3), or H0 (BF < 0.33) (Dienes, 2014).

7.3.1 Methods

7.3.1.1 Participants

52 participants (37 female, 15 male) were recruited from the NTU student population in exchange for course credits, or via an NTU student studies recruitment telegram group in exchange for $5 for every half hour of their participation. Two female participants were excluded from analysis, one due to not being in the target demographic, and one due to a loss of data following technical issues during the study. The remaining 50 participants were aged between 19 – 33 years of age (Median = 22, *SD* = 2.5). All participants reported being English-Mandarin bilingual, and having grown up mostly in Singapore. The study was reviewed by the ethics board of the university (NTU-IRB-2019-01-034) and all participants gave informed consent prior to taking part in the study tasks.

7.3.2 Procedure

The procedure for the main study was identical to that of the pilot study with the exception that participants also completed a Mandarin-specific Language ID. To collect data on our participants’ Mandarin background, they were asked to complete a series of questions. Participants were asked how old they were when they first started understanding Mandarin, how many years they had spent formally studying Mandarin (i.e., preschool, middle-high school, university), and how old they were when they first began formally studying Mandarin. They were also asked to rate the percentage of media they typically consume in local Singapore Mandarin, and standard Beijing Mandarin. Finally, they were asked to rate the percentage of time they spent communicating with others in local Singapore Mandarin, and standard Beijing Mandarin (see Appendix 7.1).

7.3.3 Analysis Plan

Decision gradient slope values were obtained for each participant in the same manner as in the pilot study. As the slope values obtained in this study did not fit a normal distribution, a log-transform was carried out on all slope values to approximate Gaussian distribution using the *log1p* function from the *base* package (R Core Team, 2020).

7.3.3.1 Pre-training control analysis

To ensure that exposure to the sound tokens presented prior to the bimodal distributional training phase had not led to any distributional learning effects, we first conducted a control analysis on the slope values of the first half and second half of the pre-training blocks. Individual slope values were obtained separately for each participant for the first half and second half of the pre-training test blocks for comparison. A paired-samples t-test was conducted to compare changes in slope values between the first half and second half of the pre-training blocks using the *t.test* function from the *stats* package (R Core Team, 2020).

7.3.3.2 Distributional learning analysis

Following this, we conducted a main analysis to determine if overall learning had occurred following bimodal distributional training, and to investigate the effects of Mandarin understanding ability on individual differences in our participants' distributional learning effects. A linear mixed effects model analysis was conducted in R (R Core Team, 2020) with the *lme* function from the *nlme* package (Pinheiro et al. 2022) on the log-transformed decision gradient slope values of all participants. The main analysis looked at the categorical fixed factor of test phase (pre-training, post-training), and the interaction between test phase and Mandarin understanding scores, with random intercept of participant, and random by-participant slope for test phase.

We expected to find a significant effect of test phase with an increase in slope values at post-training as compared to pre-training, indicating the presence of an overall distributional learning effect. We also expected to find a significant interaction between test phase and Mandarin understanding scores, signifying the presence of a language-specific transfer effect. Individual Mandarin understanding ability scores were obtained directly from the Language ID based on the number rating (from 0 – 100) that each participant gave for their Mandarin understanding ability.

7.3.3.3 EFA of Learning Microstructure and Language Factors (supplementary)

We then conducted an exploratory analysis to determine if different aspects of language background were related to distributional learning effects at different points across the study. We conducted an EFA on 14 language and learning factors using the *princomp* function (R Core Team, 2020). For this EFA, we included three language factors with scores derived from the Language ID. The three factors are as follows 1) Mandarin understanding score of each participant, 2) the total number of languages understood by each participant, and 3) the overall language understanding score of each participant.

We also included seven Mandarin background factors with scores derived from the Mandarin-specific Language ID. The seven factors are as follows 1) age of first understanding, 2) first formal education age, 3) years of formal education, 4) percentage of total media consumption in local Singapore Mandarin, 5) percentage of total media consumption in standard Beijing media, 6) percentage of total communication in local Singapore Mandarin, 7) percentage of total communication in standard Beijing Mandarin.

Finally, we included four additional learning microstructure variables derived from the decision gradient slope values at pre-test and post-test, as well as from individual test blocks in the distributional learning task: ‘overall learning’, and three variables previously investigated in Chapter 2: ‘no-training baseline’, ‘plasticity’, and ‘elasticity’.

7.3.3.3.1 Language factors

1) Mandarin understanding: The Mandarin understanding score of each participant was obtained directly from the number rating that each participant gave for Mandarin on the Language ID.
2) Total number of languages understood: The total number of languages understood for each participant was computed as the total number of languages the participant rated at 1 or higher on the language ID.
3) Overall language understanding score: The overall language understanding score of each participant was computed as the total of the number ratings each participant gave on their Language ID. For instance, a participant who rated their English understanding at 80, their Mandarin understanding at 65, and their Malay understanding at 25 would have an overall language understanding score of 170.

7.3.3.3.2 Mandarin background factors

1) Age of first understanding: The age at which each participant first started understanding Mandarin.

2) First formal education age: The age at which each participant first began formal education in Mandarin.
3) Years of formal education: Total number of years each participant spent in formal Mandarin education.
4) Percentage of local Singapore Mandarin media: Percentage of media typically consumed in local Singapore Mandarin.
5) Percentage of standard Beijing Mandarin media: Percentage of media typically consumed in standard Beijing Mandarin.
6) Percentage of local Singapore Mandarin communication: Percentage of daily conversations typically conducted in local Singapore Mandarin.
7) Percentage of standard Beijing Mandarin communication: Percentage of daily conversations typically conducted in standard Beijing Mandarin.

All scores and ages were obtained directly from each participants' responses on the Mandarin-specific Language ID.

7.3.3.3.3 Learning microstructure variables

1) "Overall learning": This variable was obtained as a measure of the overall distributional learning effect of each participant following the bimodal distributional learning phase. The variable was obtained for each participant by deducting their overall decision gradient slope value at pre-training test phase from their overall decision slope value at post-training test phase. Positive values would indicate an overall improvement in perception following training.

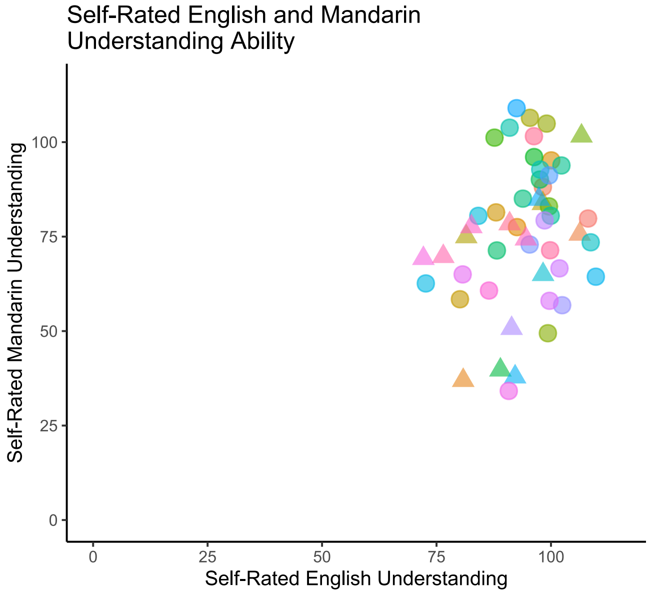
2) “No-training baseline”: This variable was obtained to determine if any participants showed an increase in sensitivity to the Singapore Mandarin alveolar-retroflex [tsʰuə́n] -[ʈʂʰuə́n] contrast continuum prior to distributional training. The variable was obtained for each participant by deducting their decision gradient slope value derived from the first half of the pre-training test phase from their decision slope value derived from the second half of the pre-training test phase. Positive values would indicate an improvement in perception prior to training.
3) “Plasticity”: This variable was obtained to determine if participants showed a distributional learning effect immediately after training, relative to their pre-training response patterns. This variable was obtained for each participant by deducting their decision gradient slope value derived from the second half of the pre-training test phase from the first half of the post-training test phase. Positive values would indicate an improvement in perception immediately following training.
4) “Elasticity”: This variable was obtained to determine if learning attrition occurred over time following the end of the training phase. This variable was obtained for each participant by deducting their decision gradient slope value derived from the first half of the post-training test block from the decision gradient slope value derived from the second half of the post-training test block. Negative values would indicate that learning attrition had occurred over time.

7.3.4 Results

7.3.4.1 Language ID

Figure 7.5

Participants' self-reported English and Mandarin understanding ability.



Note. *N* = 50. Responses jittered by 10 for visualisation.

Figure 7.5 shows the distribution of our participants' self-rated English and Mandarin understanding abilities. Overall, our participants reported high levels of understanding abilities for spoken English (Range: 80-100, *Median* = 100, *SD* = 6.3), indicating that they would not have had any issues in understanding the task instructions. Participants self-reported more varying levels of understanding abilities for spoken Mandarin (Range: 30-100, *Mean* = 75.5, *SD* = 18).

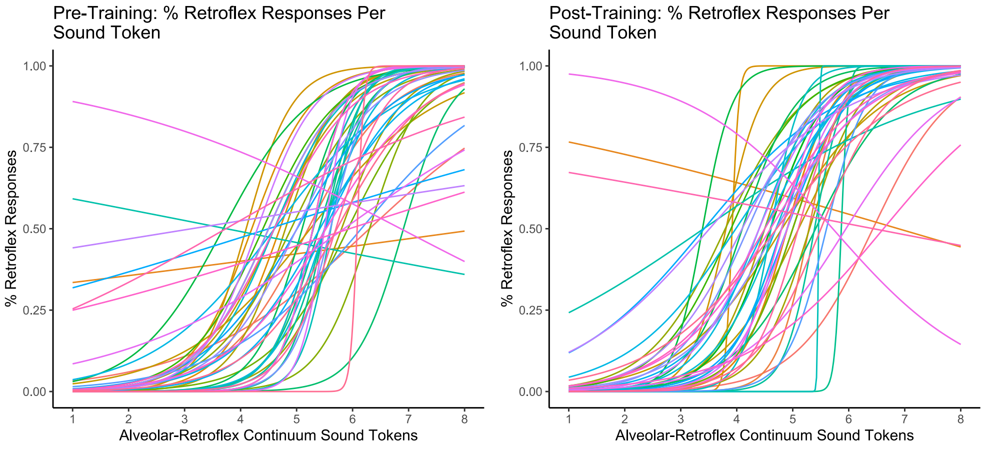
7.3.4.2 Pre-training control results

The control analysis revealed a significant difference in decision gradient slope value between the first and second half of the pre-training phase, albeit due to a significant overall *decrease* in slope values in the second half of the pre-training phase, (*t*(49) = 2.75; *p* = .008; Pre-1: *M* = .51, *SD* = .23; pre-3: *M* = .46, *SD* = .23). This indicates that no learning had occurred prior to the bimodal distributional training phase.

7.3.4.3 Distributional learning results

Figure 7.6

Individual decision gradient slopes showing fitted psychometric functions for responses at pre-training test and post-training test.



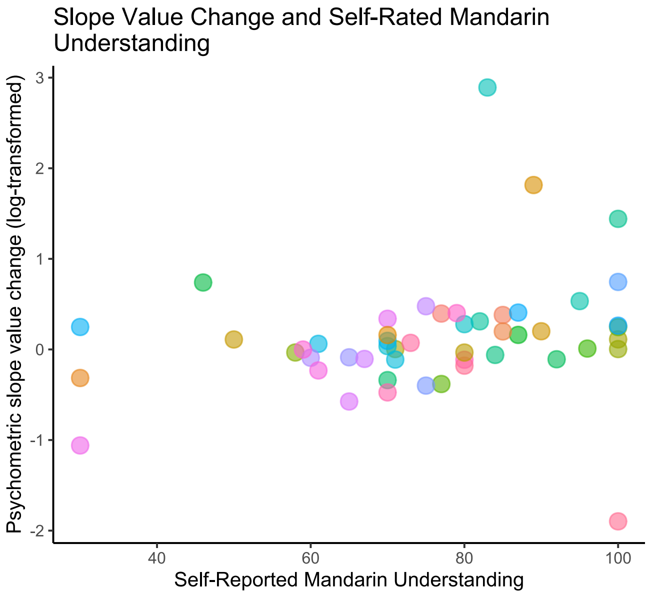
Note. *Note*. *N* = 50. Pre-training test: left. Post-training test: right.

Fitted psychometric slopes for participants’ pre- and post-training responses can be seen in Figure 7.6. Similar to the pilot study, a number of participants show visibly steeper decision gradient slopes at post-training as compared to pre-training test block. However, three participants also appear to have had no visible distributional learning effect.

The linear mixed effects model analysis conducted on the decision gradient slope values revealed that there was a significant main effect of test phase, $t(1,48) = -2.28$, $p = .03$; $\eta p^2 = .04$; $B_{H(0,.057)} = .09$; Pre-training slope values: $M = .93$, $SD = .53$; Post-training slope values: $M = 1.06$, $SD = .80$. This indicates that there was a significant overall increase in decision gradient slope values following bimodal distributional training on a Singapore Mandarin alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast.

Figure 7.7

Individual post-test minus pre-test slope differences plotted against participants' self-rated Mandarin understanding abilities.



Note. $N = 50$. Mandarin understanding abilities jittered by 2 for visualisation.

Figure 7.7 shows individual changes between pre- and post-training slope value for each participant, plotted against their self-rated Mandarin understanding ability. The linear mixed effects model analysis conducted on the decision gradient slope values also revealed that there was a significant interaction effect between test phase and self-rated Mandarin understanding ability, with larger changes correlating with higher Mandarin understanding abilities $t(1,48) = 3.09$, $p = .003$; $\eta p^2 = .17$; $B_{H(0,.0007)} = 54.6$.

7.3.4.4 EFA of Learning Microstructure and Language Factors (supplementary)

7.3.4.4.1 Language factors.

1) Refer to the Language ID section of the main study results for participants’ Mandarin understanding scores.
2) Participants reported understanding an average of 5.2 languages each (*SD* = 1.7, Range: 2-8). Note that participants typically gave understanding ratings of 50 and above only for English, Mandarin, and a third language, with fourth languages and beyond typically receiving low understanding ratings.
3) Participants had an average overall understanding score of 251.2 (*SD* = 81.4, Range: 130.0-550.0).

7.3.4.4.2 Mandarin background factors.

1) First understanding age: Participants reported an average age of 2.72 years at which they first started understanding Mandarin (*SD* = 1.23, Range: 1.0-5.0)
2) First formal education age: Participants reported an average age of 5.24 years at which they first started formal education in Mandarin (*SD* = 2.24, Range: 2.0-16.0)
3) Years of formal education: Participants reported an average of 12.7 years spent in formal Mandarin education (*SD* = 3.04, Range: 4.0-19.0)
4) Percentage of local Singapore Mandarin media: Participants reported that on average, 16.0% of the media they consumed was in local Singapore Mandarin (*SD* = 16.1, Range: 0%-61%).
5) Percentage of standard Beijing Mandarin media: Participants reported that on average, 15.1% of the media they consumed was in standard Beijing Mandarin (*SD* = 17.7, Range: 0%-82%).
6) Percentage of local Singapore Mandarin communication: Participants reported that on average, 25.8% of the daily conversations they had were conducted in local Singapore Mandarin (*SD* = 15.7%, Range: 3%-60%)
7) Percentage of standard Beijing Mandarin communication: Participants reported that on average, 5.74% of the daily conversations they had were conducted in standard Beijing Mandarin (*SD* = 10.5, Range: 0%-40%).

7.3.4.4.3 Learning microstructure variables.

1) Participants had an average ‘overall learning’ slope change of 2.03 (*SD* = 12.0, range: -15.3-81.42)
2) Participants had an average ‘no-training baseline’ slope change of -.10 (*SD* =.28, range: -1.31-.41)
3) Participants had an average ‘plasticity’ slope change of -.02 (*SD* = .20, range: -.43-.46)
4) Participants had an average ‘elasticity’ slope change of .00 (*SD* = .17, range: -.39-.42)

7.3.4.4.4 Factor loadings for EFA of Learning Microstructure and Language Factors

Table 7.1

Factor loadings on EFA of language factors and learning microstructure.

	Component					
	*1**	*2**	*3**	4	5	6
Eigenvalues	*3.17**	*2.05**	*1.48**	1.36	1.15	.946
Percent Variance Explained	*22.6%*	*14.7%*	*10.5%*	9.74%	8.19%	6.76%
Factor loadings						
Overall Learning	**.266**	.097	**.383**	.307	.592	.028
No-Training Baseline	.112	-.135	**-.567**	.382	.153	-.443
Plasticity	-.009	-.222	**.543**	.034	.430	.083

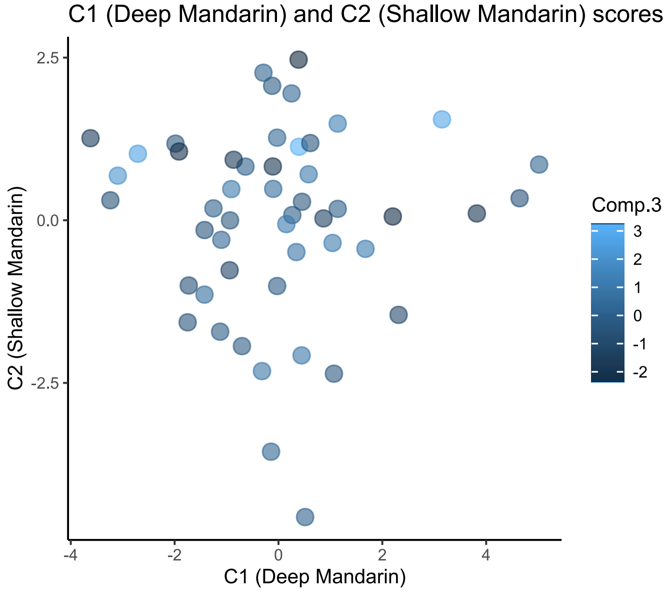
Elasticity	-.205	-.183	**-.652**	.232	.448	.192
Mandarin Understanding Ability	**.733**	**-.417**	-.132	.189	-.141	-.021
Age of Mandarin Understanding	**-.347**	**-.367**	.154	-.069	.397	-.478
Age of First Formal Mandarin Education	**-.402**	**-.414**	-.245	-.097	.189	.625
No. Years Formal Mandarin Education	**.556**	**.357**	.190	.202	.105	.227
% Media Local Singapore Mandarin	.098	**-.776**	.193	.064	-.146	.036
% Media standard Beijing Mandarin	**.642**	-.132	-.032	-.634	.189	-.103
% Communication Local Singapore Mandarin	**.278**	**-.725**	.193	.333	-.263	.061
% Communication standard Beijing Mandarin	**.589**	-.207	-.238	-.616	.175	.034
Total number of languages understood	**.581**	**.385**	-.108	.229	.068	.130

Overall language understanding score	**.855**	.027	-.082	.173	.004	.007

Note. Elasticity is inversely represented, with positive value indicating lower Elasticity. * Indicates principal components of interest identified via scree plot by identifying the ‘elbow’ in the distribution of the components (scree plot: https://osf.io/hj3kt). Factor loadings in **bold** indicate loadings of interest.

Figure 7.8

EFA Component 1 (Deep Mandarin) and Component 2 (Shallow Mandarin) coloured by Component 3 (Microlearning).



Note. $N = 50$.

Table 7.1 shows the factor loading for each component derived from the EFA, and Figure 7.8 shows the distribution of scores of Components 1 and 2, coloured by Component 3. According to Holland (2021), factor loadings of interest can be identified by the amount of information they contribute to each component. ‘Large’ loadings can be estimated as the square root of 1/number of variables per component. We chose to include Overall Learning as a factor of interest in Component 1 as it was very close to the estimated value of a ‘large’ loading (.267).

The results of the EFA conducted on the 14 language and learning factors revealed three components of interest – C1 (Deep Mandarin), C2 (Shallow Mandarin), C3 (Microlearning). Figure 7.8 shows the distribution of scores for C1 (Deep Mandarin) and C2 (Shallow Mandarin), coloured by C3 (Microlearning). Component 1 (Deep Mandarin) shows a cluster of variables related to early and successful learning of Mandarin, together with high levels of Mandarin input (particularly Beijing input) from media and in communication, as well as strong overall language scores, clustering together with overall distributional learning effects. Component 2 (Shallow Mandarin) shows a cluster of variables related to early exposure to Mandarin, together with low levels of passive Mandarin input (particularly local Singapore Mandarin media), and low active Mandarin use (local Singapore Mandarin in communication), clustering together with lower self-rated proficiency in Mandarin. This component is unrelated to distributional learning effects in the current sample. Component 3 (Microlearning) shows a cluster of variables whereby participants who showed the greatest overall learning showed a tendency towards less learning in the pre-training period (no training baseline), more change immediately after training (plasticity), coupled with a negative elasticity (i.e., a change back towards pre-training levels), meaning that their learning gains were at their peak shortly after the training period.

7.3.5 Discussion

As we predicted, our large-scale main study revealed a significant overall distributional learning effect, demonstrating that Singapore English-Mandarin bilingual adults can indeed show distributional learning effects following training on a bimodal distribution frequency of sounds along a Singapore Mandarin alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast continuum. This result is consistent with existing studies in which distributional learning effects have been successfully obtained in adults with the use of training stimuli that share featural similarities with languages in the participants' language repertoires (e.g., Escudero et al., 2011; Chládková et al., 2022). Our result of a significant overall distributional learning effect therefore lends further support to studies which have shown that distributional learning for language acquisition can be accessed well into adulthood. In particular, our results show that adults can be adept at extracting information from distributional linguistic cues when there is sufficient overlap between the featural characteristics of the training stimuli, and the speech sounds in their existing language repertoire. In this case, sharpening the discriminant function for two known words in the one of the speakers' languages on the basis of distributional information provided during training.

At the same time, the results of our study also revealed that there is indeed a language-specific component to the distributional learning effects that we found. In our exploration of a language-specific transfer effect, we predicted that our participants’ Mandarin understanding abilities would have a significant positive relationship with distributional learning effects on a distributional training paradigm with a Singapore Mandarin alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast continuum. We reasoned that this would be so, as individuals with higher Mandarin understanding proficiencies would be more likely to be exposed to Mandarin retroflexion across the lifespan through a variety of avenues such as Singapore Chinese news on television and the radio (which typically use the more ‘formal’ variety of Standard Mandarin), as well as mass media from Mainland China. As such, higher levels of Mandarin understanding proficiency are likely to be related to greater familiarity with the feature of Mandarin retroflexion, which would lead to larger language-specific transfer effects for the Singapore Mandarin alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast continuum. This prediction was indeed borne out in the significant interaction effect that we found. Visual inspection of Figure 7.7 which shows a clear effect of higher Mandarin understanding abilities being associated with larger distributional learning effects. Moreover, the effect size that we obtained for this interaction was large (ηp^2 = .17), and we obtained a Bayes factor of 54.6 for this interaction. This Bayes factor indicates that the data collected were 54.6 times more supportive of a speaker’s Mandarin understanding impacting the strength of the observed distributional learning effect, rather than having no impact. Thus, we can state with a high degree of confidence that we have obtained evidence of a language-specific learning advantage in our Singapore English-Mandarin bilingual adults, and this effect is stronger for bilinguals with stronger language skills in the target language. This is consistent with existing studies which show that language learning effects are typically enhanced by featural and structural overlaps between participants' languages and the training stimuli (e.g., Antoniou et al., 2015). Crucially, while existing studies on language-specific learning effects in adults have typically compared different groups of native language speakers (e.g., Greek vs Czech speakers; Chládková and Šimáčková, 2021) our study shows that language-specific transfer effects can be strong enough to be observed even within the individual differences in language backgrounds for a single group of Singapore English-Mandarin bilingual adults. Moreover, as the results of our earlier perception study in Chapter 5 revealed that baseline levels of contrast sensitivity for this synthesised Singapore Mandarin alveolar-retroflex contrast are *not* in fact linked to overall Mandarin understanding abilities, the interaction effect that we found here appears to show that there may be a specific transfer advantage linked to Mandarin exposure regardless of baseline sensitivity to these sounds.

However, the Bayes factor that we obtained for the main overall distributional learning effect was unexpectedly substantial in the opposite direction that we predicted. The Bayes factor we obtained for the main distributional learning effect was BF = .09. As this is less than .33, this appears to show that our data are substantially supportive of the null hypothesis, i.e., that there was no overall distributional learning effect following bimodal distributional training. Therefore, it is important to note that we obtained a significant main effect of test phase in the presence of a significant interaction between test phase and Mandarin understanding ability (with a large interaction effect size). As such, the distributional learning effects obtained in our main study would naturally have been moderated by the range of individual differences in our participants' Mandarin understanding abilities. As this resulted in a smaller overall distributional learning effect in our larger sample size as compared to in the pilot study, the Bayes factor we obtained could be obscuring a very real distributional learning effect. Schmalz et al. (2023) found that false negatives in Bayesian hypothesis testing can be very common (up to 80% of the time) if an effect size found is much smaller than the chosen prior. Moreover, based on the a priori analysis that we conducted to determine a well powered sample size, a sample size of $N = 48$ would have been sufficient to obtain a similar distributional learning effect to that of the pilot study. Thus, we can conclude that there was likely a modest overall distributional learning effect in our main study which cannot be interpreted on its own without accounting for the highly significant interaction effect between distributional learning and Mandarin understanding that we also found.

Our supplementary EFA also revealed interesting relationships between our participants' distributional learning effects, and various aspects of their linguistic and Mandarin background. C1 (Deep Mandarin), shows that stronger overall distributional learning effects are linked to not only to Mandarin understanding abilities and early language exposure, but also to the degree of engagement that individuals have with Mandarin in daily life (i.e., in communication and media consumption. This shows that high levels of daily active and passive engagement with Mandarin may facilitate the language-specific transfer effects seen in the study. At the same time, C1 (Deep Mandarin) also showed that high levels of active and passive exposure to *Beijing* Mandarin had a larger impact on overall distributional learning effects as compared to the local variety of Mandarin. This could indicate that exposure to the clear alveolar-retroflex contrast present in standard Beijing Mandarin could have carry-over transfer effects for phoneme contrast sensitivity for an alveolar-retroflex contrast in Singapore Mandarin. C1 (Deep Mandarin) also showed that stronger distributional learning effects were linked to higher understanding proficiencies in a wider range of languages. This could lend some support to the domain-specific, language-general account of language learning advantages by demonstrating that the experience of successfully managing large language repertoires may be linked to metalinguistic advantages in multilinguals.

While C2 (Shallow Mandarin) was not related to any distributional learning effects, it demonstrated that earlier Mandarin exposure and longer periods of formal Mandarin education can be linked to weaker Mandarin understanding abilities in the presence of low levels of engagement with Mandarin in daily life. This suggests that early exposure and longer periods of formal education alone may not be sufficient for individuals to attain high levels of Mandarin understanding abilities. Instead, higher levels of both passive and active engagement with the language may be needed to attain high levels of Mandarin proficiency.

Finally, C3 (Microlearning) was not related to any language effects, and instead showed that some participants may exhibit learning patterns in which weaker pre-training contrast sensitivity is linked to short-lasting increases in contrast sensitivity immediately following training, but resulting in strong overall distributional learning effects. This suggests that high general baseline levels of perceptual sensitivity for this Singapore Mandarin alveolar-retroflex contrast are not needed for individuals to show strong distributional learning effects following distributional training, although this effect declines over time-since-training for some participants.

7.4 Conclusion

On the whole, this study has shown that the effects of bimodal distributional training can in fact be found well into adulthood. The results of both our pilot and main study demonstrate that our target population of adult Singapore English-Mandarin bilinguals are indeed able to show significant overall distributional learning effects following distributional training on a Singapore Mandarin alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast continuum. By comparing the distributional learning effects of this study to that of our first distributional learning study which we conducted with a completely novel

Hindi dental-retroflex contrast, we can see that a language-specific overlap between known speech sounds, and to-be-learnt language stimuli plays an important role in determining whether adult participants will even be capable of extracting linguistic information from distributional cues.

Moreover, the impact of a language-specific transfer effect can be clearly seen in the significant interaction that we found between distributional learning effects and Mandarin understanding abilities in our main study. This result revealed that participants with higher Mandarin understanding ability scores showed greater distributional learning advantages (i.e., larger increases in contrast sensitivity) for a fine-grained Singapore Mandarin alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast continuum as compared to participants with lower Mandarin understanding ability scores. As such, we can confidently say that we have found evidence for the language-specific hypothesis of bilingual learning advantages that we first set out to test at the start of this project. However, such an effect only emerges when the acoustic features of the stimuli are finely tuned to the local linguistic variety, such that a speaker can transfer information from their mental representation to the task. In the current study, we were able to achieve this through a combination of preliminary bioacoustic imaging (Chapter 3), psychoacoustic analysis (Chapter 4), and perceptual mapping of acoustic space (Chapter 5). This multidisciplinary approach provides nuanced insights into the specificity of linguistic representations for speakers of Singapore Mandarin.

Chapter 8: General Discussion and Conclusion

8.1 Overall Discussion

In this project, we set out to determine if we could find evidence for a language-specific transfer advantage in bilingual adults. According to this language learning hypothesis, language-learning advantages are specific to the individual's linguistic experience, and learning advantages are closely tied to featural and structural overlaps between the individuals' known languages, and the to-be-learnt language stimuli. We chose to focus our investigation on contemporary Singapore English-Mandarin bilingual adults as they consist of a group with a largely homogenous experience of language exposure to English and Mandarin through daily life and formal education, albeit with intragroup heterogeneity in their degree of proficiency with Mandarin Chinese. In line with the language-specific hypothesis of learning advantages, we reasoned that individuals with a higher proficiency in Mandarin Chinese would have an advantage at learning language stimuli that share featural regularities with Mandarin Chinese, but not with English. We chose to investigate this with a distributional learning paradigm as studies have shown that it is a robust method of identifying bilingual learning advantages tied to language experience in adults (Chládková & Šimáčková, 2021; Chládková et al. 2022). Moreover, distributional learning paradigms are easily adapted for participants across the lifespan, for example with younger participants (e.g., Infants: Maye et al., 2022; Children: Vandermosten et al., 2018).

In our first study, we explored the language-specific hypothesis with a methodological adaptation of a distributional learning study by Vandermosten et al. (2018). We trained English-Mandarin bilingual adult participants on a bimodal distributional frequency of sounds drawn from a completely novel 7-step Hindi dental-retroflex /d̪a/-/ɖa/ contrast continuum, and tested their sensitivity for this contrast before and after the distributional training. As Hindi and Mandarin Chinese both ostensibly feature a contrast between alveolar/dental and retroflex places of articulation (while English does not), we reasoned that individuals with higher Mandarin understanding abilities should show larger distributional learning effects for this Hindi contrast. However, we did not manage to find an overall distributional learning effect for this Hindi dental-retroflex /d̪a/-/ɖa/ contrast, and were unable to observe graded language-specific learning effect in our English-Mandarin bilingual adult participants. The lack of an overall distributional learning effect was unlike that of Vandermosten et al. (2018) who reported significant overall learning results in their child participants. It is possible that the stimuli used may have been too difficult for our adult participants to learn with a distributional learning paradigm. Indeed, studies have suggested that adults may face particular difficulties in extracting distributional cues from novel language stimuli (e.g., Wanrooij et al., 2014). Moreover, research published after the study in Chapter 2 was completed has since revealed that there may not be featural regularities shared between the kind of retroflexion in Hindi and contemporary Mandarin phonology, as the Mandarin “retroflex” has a tongue posture that differs from the “curled” retroflexion found in Hindi (Taiwan Mandarin: Chiu et al., 2020; Tiede et al., 2019; Standard Mainland Mandarin: Luo, 2020). To investigate whether this was also the case in the local Singapore variety of Mandarin, we conducted a preliminary observational lingual ultrasound study in Chapter 3. This study found that none of the participants that we reported on showed evidence of a “curled” tongue position in their production of Singapore Mandarin retroflexion. This suggests that the Hindi dental-retroflex contrast continuum used in Chapter 2 was not well-suited for eliciting distributional learning via transfer in Singapore English-Mandarin bilingual participants. Note that existing examples of successful distributional learning in adults largely involve the use of stimuli that contain structural overlaps with the participants’ languages (e.g., Chládková & Šimáčková, 2021; Chládková et al. 2022; Escudero et al., 2011). Chapter 2 reveals that we would not be able to effectively assess language-specific transfer effects due to the lack of any featural or structural overlaps between the training stimuli and the Mandarin alveolar-retroflex contrast.

Thus, it was evident at this point that a locally valid set of language stimuli would be needed to accurately assess the language-specific hypothesis of bilingual learning advantages in our target population. However, little research into contemporary Singapore Mandarin was available at this point for such purposes. As such, we conducted a large-scale perception and production study on the alveolar-retroflex contrast in Singapore Mandarin in Chapter 4 to obtain the acoustic documentation needed for our investigation to continue. We utilised CoG value measurements and GAMM on LTAS data to measure alveolar and retroflex frication production in Singapore Mandarin for the first time in known literature. From our investigation, we discovered that contemporary Singapore Mandarin speakers generally exhibit high levels of Mandarin alveolar-retroflex word identification accuracy in both the local as well as the standard Beijing varieties of Mandarin. This contrasts with findings from Taiwan Mandarin speakers who have been documented to show poorer contrast perception, likely due to the influence of deretroflexion in Taiwan Mandarin (Shih & Kong, 2011). At the same time, we found that contemporary Singapore Mandarin speakers do indeed produce word-initial alveolar and retroflex phonemes contrastively (i.e., the acoustics of the productions are significantly different), albeit with a range of individual differences in the degree of contrast produced. This is different from earlier observational studies of Singapore Mandarin in the 1980s which have suggested almost complete merger of the alveolar-retroflex contrast in speech (Lock, 1989; Ng, 1985). This suggests that modern influences, such as the formal education system and Mainland Mandarin media may be playing a role in influencing the phonology of contemporary Singapore Mandarin.

Based on the corpus of contemporary Singapore Mandarin speech and the acoustic measurements that we obtained in Chapter 4, we then synthesised a novel set of language stimuli consisting of an 8-step alveolar-retroflex contrast continuum based on the acoustic norms of the words [tsʰuə́n]-[ʈʂʰuə́n] (村

‘village’ - 春 ‘spring’) as produced by our Singapore Mandarin speakers. Our goal was to use this set of stimuli to further investigate the language-specific hypothesis of bilingual learning advantages. Thus, we first needed to determine if this set of stimuli could be effectively used to capture fine-grained measures of individual differences in perceptual sensitivity for this contrast in the local target population. We designed a novel 2AFC task based on the CROWN Game of Ke et al. (2021) – the Spring Village CROWN Game – designed to measure ambiguity resolution for sound tokens on the 8-step [tsʰuə́n]-[ʈʂʰuə́n] continuum. We validated the Spring-Village CROWN Game in a sample of our target population of Singapore English-Mandarin bilingual adults in Chapter 5. The results of our study in Chapter 5 showed that the majority of contemporary Singapore Mandarin speakers show evidence of perceiving a clear contrast between alveolar and retroflex places of articulation, in line with the word identification task results of Chapter 4. At the same time, the results also demonstrated that the Spring-Village CROWN Game was capable of effectively capturing fine-grained individual differences in perceptual sensitivity for different degrees of acoustic ambiguity embedded in the [tsʰuə́n]-[ʈʂʰuə́n] alveolar-retroflex contrast continuum. Thus, we had good reason to believe that this would be well-suited for adaptation for a distributional learning task that would allow us to examine language-specific transfer effects amongst Singapore English-Mandarin bilingual adults in a manner that was valid to the variety of Mandarin spoken in contemporary Singapore.

Before we began this final distributional learning study, we conducted an exploratory study in Chapter 6 using methods we had established so far to determine if we could find perception-production links in Singapore Mandarin speech for a [tsʰuə́n]-[ʈʂʰuə́n] alveolar-retroflex contrast. We conducted this study as research has shown that such links for speech sounds may exist on the level of the individual (Fox, 1982; Hearnshaw et al., 2018; Newman, 2003). Moreover, studies have found that perceptual sensitivity training on speech sounds can result in concomitant changes in speech production (Bradlow et al., 1997; Inceoglu, 2016; Sakai & Moorman, 2018). Evidence of perception-production links would thus give us future directions to consider following our final distributional learning study. The results of this study showed that there is a significant perception-production link tied to the retroflex place of articulation in Singapore Mandarin. Participants who produced retroflex frication that was more similar to typical alveolar frication in acoustic space (deretroflexion) were found to show less perceptual sensitivity for retroflex cues in ambiguous speech sounds. This is aligned with existing studies on second language acquisition and articulatory disorders which show that individuals typically have reduced perceptual sensitivity for speech sounds that they produce less clearly (e.g., Rvachew & Jamieson, 1989; Sakai & Moorman, 2018).

Finally, we completed our investigation into the language-specific hypothesis of bilingual learning advantages with a distributional learning study in Chapter 7. We trained English-Mandarin bilingual adult participants on a bimodal distributional frequency of sounds drawn from our synthesised 8-step alveolar-retroflex [tsʰuə́n]-[ʈʂʰuə́n] contrast continuum, and tested individual differences in sensitivity for this contrast before and after the distributional training with the Spring-Village CROWN Game. The contrast continuum used in this study was synthesised to be acoustically aligned to the norms of alveolar-retroflex frication produced by contemporary Singapore Mandarin speakers. Thus, based on the language-specific hypothesis of bilingual learning advantages, we reasoned that individuals with higher Mandarin understanding abilities should show larger distributional learning effects due to a larger degree of overlap between their language repertoires and the training stimuli. The results of the main study of Chapter 7 revealed that our participants showed a significant overall distributional learning effect following bimodal distributional training – this is consistent with studies which show that adults *can* successfully extract distributional cues from training stimuli (Chládková & Šimáčková, 2021; Chládková et al. 2022; Escudero et al., 2011; Escudero & Williams, 2014; Wanrooij et al., 2013).

Moreover, we found a significant positive interaction between distributional learning effects and Mandarin understanding abilities. As we predicted, participants with higher Mandarin understanding abilities were found to have larger distributional learning effects. This supports previous studies that have shown that bilinguals have language-specific transfer abilities tied to their language repertoires, where language experience facilitates the transfer of linguistic knowledge between known languages and to-be-learnt language stimuli (e.g., Antoniou et al., 2015; Chládková & Šimáčková, 2021; Chládková et al. 2022; Kuo et al., 2016). In particular, while evidence for language-specific theories have largely been found in the context of comparing monolinguals to bilinguals, or by comparing different groups of bilingual speakers (e.g., Antoniou et al., 2015; Bialystok et al., 2003; Kuo et al., 2016) our study has shown that language-specific transfer effects are robust enough to be identified within individual differences in the language proficiencies of a largely homogenous group of bilingual adults.

8.2 Discussion of the language-specific account of learning advantages in Singapore English-Mandarin bilingual adults

8.2.1 Implications

The nature of language learning advantages can largely be classified into three main overarching hypotheses: domain-general accounts suggesting that bilinguals have general cognitive advantages, domain-specific and language-general accounts suggesting that bilinguals have general metalinguistic advantages, language-specific accounts suggesting that language learning advantages in bilinguals are tied to their unique linguistic experience. In this project, we have found evidence to support language-specific accounts of learning advantages in bilinguals. In Chapter 7, we found that English-Mandarin bilingual adults showed larger distributional learning effects for a novel synthesised [tsʰuə́n]-[ʈʂʰuə́n] alveolar-retroflex contrast if they had higher Mandarin understanding proficiencies. This shows that larger transfer effects were found for participants who had more extensive experience with a relevant language. Moreover, the supplementary EFA that we conducted on various aspects of our participants' Mandarin experience also shows that familiarity with both the local variety as well as a different variety of Mandarin (standard Beijing Mandarin) are linked to larger distributional learning effects. This suggests that language-specific metalinguistic advantages in bilinguals may be enhanced by experience with any relevant languages and varieties that share featural regularities with the to-be-learnt stimuli. As such, it is possible that experience with local vernacular languages may facilitate the learning of "official" languages in the formal education system. Local vernacular languages often share large degrees of featural and structural regularities with the "official" languages. For instance, studies have suggested that the large degree of phonological overlap between Hokkien and Mandarin may give Hokkien speakers an advantage at learning Mandarin phonology (Newman, 1982; 1983). Evidence for a cross-language transfer effect between Hokkien and Mandarin has also been found by Newman (1982) who found that fluent Hokkien speakers were successfully able to transfer their knowledge of tones to Mandarin speech with a high level of accuracy. This demonstrates that local home languages from the Chinese language family could potentially be beneficial for the acquisition of 'Mother Tongue' languages in formal education, particularly if there are phonological overlaps shared between the languages. This finding also aligns with a previous report of young adults' self-assessed language proficiency and early childhood language exposure, in which earlier and larger exposure to non-Mandarin Chinese varieties was associated with advantages in Mandarin proficiency in later life (Wu et al., 2020). This suggests that the use of local vernacular home languages may, in fact, be mutually supportive of the acquisition of ‘Mother Tongue’ languages. The use of these home languages may be encouraged to facilitate acquisition of ‘Mother Tongue’ languages.

8.2.2 Limitations

To date we have only examined language-specific transfer effects in a single group of bilinguals. This limits the generalisability of the results of our study. To more clearly determine if the learning effect we found is *specific* to the language experiences of Singapore English-Mandarin bilingual adults, it would be of value to conduct a replication of our study in Chapter 7 with a different group of bilinguals with language repertoires that do not share phonological regularities with the current set of stimuli. A follow-up study can be conducted with Singapore English-Malay bilinguals to achieve this – data collection for this follow-up study has already begun.

Next, as we focused our investigation solely on bilinguals, the results of our study do not tell us if there is also a language-general, domain-specific metalinguistic advantage that comes solely from the experience of being bilingual or multilingual. In order to determine if there is indeed a language-general metalinguistic learning advantage tied to the experience of bilingualism, one possibility would be to conduct a follow-up study on a well-matched comparison group of monolingual Mandarin speakers. This would be difficult to achieve in the context of Singapore as the majority of residents in the country are either bilingual or multilingual (DOS, 2021), and almost all young adults who report monolingualism are monolingual in English. Monolingual participants from other countries could be considered for such a follow-up study. However, differences in the varieties of Mandarin spoken in different countries would be likely to have a confounding effect on the results which would make comparisons with the current study difficult to interpret.

A monolingual comparison group would undoubtedly have interesting implications for determining if the experience of acquiring multiple languages might confer additional learning advantages on bilinguals. However, existing research comparing learning between monolingual and early bilingual adults has typically found mixed results. It is thus possible to hypothesise that we may obtain a variety of different outcomes with a comparison group of Mandarin monolinguals for the distributional learning study in Chapter 7.

First, it is possible that English-Mandarin bilinguals may demonstrate an overall distributional learning advantage as compared to a well-matched Mandarin monolingual group of participants. This would be similar to the results of Antoniou et al. (2015) who trained English monolingual and English-Mandarin early bilingual adults on synthesised English-based frication contrasts, and found that the bilingual group showed larger learning effects as compared to the monolingual group. Alternatively, such a study might find that different subsets of English-Mandarin monolinguals may perform differently against Mandarin monolinguals depending on whether they are English-dominant or Mandarin-dominant in their daily language use. This would be similar to the results of a different study by Antoniou et al. (2012) who found that English-dominant Greek-English early bilingual adults performed as well as English-monolinguals at learning English-based contrasts, but were outperformed by Greek monolinguals at learning Greek-based contrasts.

In light of these different possible outcomes, is important to highlight that even if we were hypothetically able to obtain a comparison group of monolingual contemporary Singapore Mandarin speakers to examine language-specific transfer effects as in Chapter 7, this comparison group would have to be well-matched to the Mandarin proficiencies of contemporary Singapore English-Mandarin bilinguals to avoid a confounding effect of higher overall proficiencies in Mandarin-only monolinguals. The acquisition of such a Mandarin monolingual comparison group would come with its own set of challenges. It would be difficult to ensure that monolingual Mandarin-speaking participants with low Mandarin proficiencies would be able to understand and perform the task as required, and it would also be necessary to adequately screen for potential learning disabilities that might result in low language proficiencies in Mandarin monolinguals.

This discussion thus highlights how crucial it is to ensure that the research design at hand has *ecological validity* to the population in question. This is particularly important when conducting research on typically underdocumented populations with unique differences in language diversities.

The lack of consideration for global linguistic diversities is a widespread issue in existing literature. For instance, a review of literature by Kidd and Garcia (2002) found that the majority of papers published on child language acquisition are disproportionately focused on western monolingual English populations. Similar trends can also be seen in adult psycholinguistic research. Analyses of existing literature show that only 0.6% of languages in the world have been investigated (Jaeger & Norcliffe, 2009), with the vast majority of studies focusing on only 10 languages, and English studies taking up a large proportion of this research real estate (Anand et al., 2011). Thus, forcing the investigation to fit the framework of previous research that has typically been conducted in western, industrialised populations would be a disservice to research participants in a population that does not fit these standards and would be akin to insisting that all psycholinguistics research conducted in primarily monolingual countries should also have a bilingual comparison group in order to adequately make sense of their findings.

As such, the lack of a monolingual comparison group *could* be viewed as a limitation in terms of how closely our research conforms to the existing conventions of bilingualism research, wherein bilingualism is often viewed through the context of its differences from monolingualism. However, this thesis instead proposes a different framework of treating bilingualism as a standalone linguistic phenomenon which can and should be examined by virtue of its own relevance as the mainstream language model of linguistic repertoires in contemporary Singapore. As such, this project is primarily interested in examining bilingualism as the standard language model in Singapore by investigating how bilinguals differ amongst themselves based on individual differences in their language backgrounds. The results of this project therefore demonstrate the power of investigating bilingualism as a continuous variable within a single group of participants who share languages of exposure but differ in their language proficiencies (i.e., the depth of bilingualism in Mandarin).

8.2.3 Future directions

In Chapter 7, we largely focused our examination of language-specific transfer effects on Singapore Mandarin speakers learning a novel set of ambiguous stimuli based on the acoustic norms of Singapore Mandarin. It would be of interest to determine if we can also find evidence for a *cross-language* transfer effect between our participants' known languages, and a less-familiar set of stimuli based on a different language. Previous literature has described overlaps between Hokkien and Mandarin phonology and grammatical structural regularities (Newman, 1982; Newman 1983; Newman, 1988). Hence, a follow-up study could be conducted on Singapore English-Mandarin bilingual adults with Hokkien-based training stimuli, for example Hokkien tones. Given the transfer effects seen in Chapter 7, we predict that we would be able to find evidence for such a cross-language transfer effect. Such a study would strengthen the validity of the language-specific hypothesis of learning advantages in bilinguals by demonstrating clear evidence of a cross-language transfer effect.

8.3 Discussion of distributional learning in adults

8.3.1 Implications

We conducted two distributional learning studies (Chapter 2 and Chapter 7) to elicit learning effects for phoneme contrasts in adults. These two studies have provided us with more insight into the mechanism of distributional learning in adults, and provide more support for existing studies. Some research has suggested that adults lack the sensitivity needed to extract distributional cues from language training stimuli (Liu et al., 2022; Terry et al., 2015; Wanrooij et al., 2014). On the other hand, other studies have shown that adults *can* successfully extract distributional cues from language training stimuli (Chládková & Šimáčková, 2021; Escudero et al., 2011; Wanrooij et al., 2013), and may even show long-term learning from such training paradigms (Escudero & Williams, 2014).

One crucial difference between these two different sets of results is that the studies with adult participants who failed to show any distributional learning effects typically used training stimuli that contained no linguistic overlaps with the languages in their participants' language repertoires (e.g., Mandarin-based tone stimuli with Australian English speaking adults, Liu et al., 2002). However, the studies with adult participants who showed robust learning effects typically used training stimuli that contained some degree of linguistic overlap with the languages in the participants' existing language repertoires (e.g., durationally cued vowel contrasts with Czech speaking adults).

Our two distributional learning studies support the findings of these studies. In Chapter 2, we did not find an overall distributional learning effect in our English-Mandarin bilingual adult participants following training on a specific Hindi dental-retroflex contrast that did not share linguistic overlap with any of the languages in their language repertoire. However, we found a significant overall learning effect in Chapter 7 when our English-Mandarin bilingual adult participants were trained on a Singapore Mandarin-based alveolar-retroflex contrast. This shows that adults can in fact successfully utilise distributional learning for novel language stimuli if there is sufficient linguistic overlap between the stimuli and their known languages. Moreover, our study in Chapter 7 has shown that adults can be sensitive enough to extract distributional cues in an unattended training paradigm while focusing their attention on a different set of visual stimuli (muted cartoons). This suggests that the unattended training paradigm can be adapted to facilitate perceptual sensitivity enhancements for speech sounds in other unattended settings.

8.3.2 Limitations

The validity of the distributional learning effects found in Chapter 7 could be further clarified through the inclusion of a control group to demonstrate whether changes in perceptual sensitivity are indeed the direct result of bimodal distributional training as opposed to a test-retest effect. A control group could consist of a separate group of individuals trained on a unimodal frequency distribution of sounds drawn from the synthesised Singapore Mandarin alveolar-retroflex contrast (e.g., Chládková & Šimáčková, 2021; Wanrooij et al., 2014) or exposed to a similar duration of unrelated sounds such as music during the training phase (e.g., Escudero & Williams, 2014; Wanrooij et al., 2013).

As such, a control group exposed to a different set of stimuli during training could serve to rule this possibility out. While the in-group comparison of pre- and post-training perceptual differences has been used in existing research as a standalone analysis to determine if distributional learning effects have occurred following bimodal distributional training (e.g., Chládková et al., 2022), it is possible that changes in perception at post-training may simply be due to changes in perceptual sensitivity for these sounds following the natural passage of time. The inclusion of a well-matched control group will be of great research value in future distributional learning studies.

Despite this limitation, it is important to note that our investigation of language-specific transfer effects was primarily interested in determining if changes in perceptual sensitivity might be linked to the Mandarin proficiencies of our participants. As this is indeed what we found in Chapter 7, our results nevertheless show that there is a language-specific component linked to the increase in perceptual sensitivity seen in participants at post-training, regardless of the process that resulted in this change.

8.3.3 Future directions

As mentioned in Section 8.2.3, it would be of interest to investigate if our target population of English-Mandarin bilingual adults would be able to successfully extract distributional cues from a set of stimuli based on a language that they are less familiar with (e.g., a Hokkien-based contrast). Conducting such a study would allow us to determine the degree of familiarity that adults need to have with different language structures in order for successful distributional learning to occur. At the same time, while we have shown in Chapter 7 that adults do show learning effects following bimodal distributional training on speech sound stimuli, we do not yet know how long these learning effects can last for. It would be of interest to conduct a follow-up study in line with that of Escudero and Williams (2014) to determine if the bimodal distributional training might likewise have long-lasting effects on perceptual sensitivity.

Finally, as we have only conducted our distributional learning studies on adults, it would be of interest to determine if the results we have found can be generalised to participants at other points of the lifespan. In future follow-up studies, the distributional learning paradigm we used in Chapter 7 can be adapted for infants and children to investigate this question (e.g., with the use of eye tracking for infants and young children). Based on existing research, we predict that younger participants will be able to show robust learning effects in these studies.

8.4 Discussion of perception and production of Singapore Mandarin alveolar-retroflex sibilant contrast

8.4.1 Implications

Our studies on perception and production of the contemporary Singapore Mandarin alveolar-retroflex contrast (Chapter 4, 5, and 6) comprise the first time in which detailed documentation have been conducted on this contrast in known literature. We can see from these studies that the phonology of contemporary Singapore Mandarin spoken by young adults is different from that of adults in the 1980s. Earlier observational studies found evidence of a large degree of deretroflexion or almost complete alveolar-retroflex phoneme contrast merger in the speech of adult Singaporeans (Lock, 1989; Ng, 1985). However, our studies show that contemporary Singapore Mandarin speaking adults largely produce these phonemes contrastively in speech (Chapter 4 and Chapter 6) and typically have high levels of perceptual contrast sensitivity for word identification (Chapter 4) and speech sound ambiguity resolution (Chapter 5 and 6).

Documentation of these differences is important as these details can have implications on the variety of Mandarin that should be used in different situations in daily life, such as in formal education settings. In Chapter 4, we discussed the importance of documenting outer-circle languages to ensure that speech audiometry will be valid to the local population. Speech audiometry material are necessary in the medical field for the purposes of identifying and diagnosing hearing difficulties linked to specific phonological aspects of speech. According to Soh & Loo (2023), the use of audiometry material not aligned to local varieties can lead to inaccurate results due to listeners being inherently unfamiliar with the phonology of the stimuli. This has been identified as a particular issue in Singapore, as there is no existing audiometry material that is valid for the local context of Singapore Mandarin speakers (Soh and Loo, 2020). Our study shows that there may be considerable intergenerational difference in the phonology of Mandarin spoken across age groups in Singapore. This suggests that future work on speech audiometry material in Singapore should not only focus on ensuring that stimuli are valid to Singapore Mandarin, but that different sets of material are available for accurate speech hearing tests across different generations.

On the whole, the studies we have conducted in this project show how important it is to conduct fine-grained documentation on the specific features of local varieties of languages to ensure that speech stimuli used in studies of language learning and perception are well aligned with the acoustic norms of the target population. This project also demonstrates the effectiveness of applying interdisciplinary methods from linguistics and psychology to obtain a comprehensive understanding of the research topic at hand.

8.4.2 Limitations

It is worth noting that some of the methodological decisions in our Singapore Mandarin perception and production tasks could potentially have resulted in unintended priming effects leading to participants producing

Mandarin speech in a different manner to how they normally would in daily life. For instance, participants in Chapter 4 completed the speech production task after the word identification task. It is thus possible that there could have been some degree of accent transfer from the stimuli that they heard to their production of Mandarin speech. However, as participants heard an equal number of Singapore Mandarin and Beijing Mandarin word tokens in random order over the course of the word identification task, it is unlikely that they would have been more influenced by one variety of Mandarin as compared to the other. Moreover, the speech production task did not occur immediately after the word identification task. Instead, there was a transition period of 5 – 10 minutes between each task during which participants were given instructions in English on how to use the recording equipment and perform the speech production task. It is therefore unlikely that the participants would have retained a sufficient amount of influence from the word identification task to significantly affect their production of the Mandarin alveolar-retroflex words. Moreover, it is worth noting that a similar result of a significant difference between the production of alveolar and retroflex frication in Singapore Mandarin was also found in Chapter 7 where participants completed the speech production task *before* the word identification task.

It is also worth considering that the inclusion of Hanyu Pinyin on the alveolar-retroflex minimal word pair picture cards could have prompted our participants to be more aware of the alveolar-retroflex contrast, and to thus produce the contrast more clearly in speech. However, the use of word-reading tasks is not exclusive to this study. Reading from a list of words has been used as an established method of documenting accent varieties in English, where the spelling of English words may inherently provide participants with cues as to the pronunciation of different words. This can be seen in the documentation of English vowels by Peterson and Barney (1952) as well as in Wells Lexical Set (Wells, 1982), and even for vowels in Singapore English (Deterding, 2007). Additionally, we determined that the inclusion of Hanyu Pinyin was necessary to ensure that participants with a wide range of Mandarin proficiencies would be capable of reading the words on the cards, as individuals assisting in the design of the study reported difficulties in being able to read and recognise the words when presented as Chinese characters alone. Moreover, even with the inclusion of Hanyu Pinyin on the picture cards, the results of Chapters 4 and 6 nevertheless show that there is still a wide range of individual differences in how contrastively alveolar and retroflex frication are produced in Singapore Mandarin speech. This suggest that individual differences in production of the Singapore Mandarin alveolar-retroflex contrast exist even in the presence of cues that might highlight the contrast in question.

Besides this, it is possible that the use of a word reading task for speech data collection could have led to contrastiveness in production of the alveolar-retroflex contrast being better preserved as compared to speech production data collected from conversational free-speech prompts (e.g., Chang 2012). However, it would have been more difficult to elicit natural utterances via conversational styles of data collection as the primary researcher and data collector for this project does not speak or understand Mandarin, which could have reduced the ecological validity of a conversational prompt. Moreover, we believe that the data collected on production of the contemporary Singapore Mandarin alveolar-retroflex contrast nevertheless reflects genuine evidence supporting the increase in retroflexion in contemporary Singapore Mandarin speech in comparison to the speech of young adults in the 1980’s. This is particularly so as the word-reading methodology used in this project was also used by Ng (1985) to collect data on the use of the retroflex /ʂ/ phoneme in Singapore Mandarin speech of the time.

Finally, this project would likely have benefitted from consultation with a Mandarin linguist from the outset, as this would have allowed us to better understand the nature of “retroflexion” in Mandarin before conducting our first distributional learning study in Chapter 2. Despite this, it is important to note that the results of Chapter 2 prompted further investigation into the alveolar-retroflex contrast in Singapore Mandarin, yielding novel information on the linguistic nature of the contrast, as well as detailed acoustic documentations of the contemporary Singapore Mandarin alveolar-retroflex contrast for the first time in known literature. This information has far-reaching implications beyond that of research, which we have discussed in greater detail in Section 8.4.1.

8.4.3 Future directions

While we have collected some ultrasound data and performed a preliminary visual inspection of the lingual positions used in alveolar and retroflex frication in Singapore Mandarin in Chapter 3, we have yet to conduct a more fine-grained statistical analysis examining the precise nature of lingual articulation for the alveolar-retroflex contrast in Singapore Mandarin. At this point in time, we have collected lingual ultrasound data from 21 participants, and plan to conduct a follow-up study by conducting a principal component analysis (PCA) and general additive mixed modelling (GAMM) on the lingual ultrasound traces of the speech tokens of all 21 participants. As we have also collected lingual ultrasound data on the production of the alveolar-post-alveolar contrast in Singapore Mandarin from these 21 participants, we likewise plan to conduct similar analyses on this data to examine if there are overlaps between the Singapore Mandarin alveolar-retroflex contrast and the Singapore English alveolar-post-alveolar contrast. This will provide us with a clearer picture of the nature of lingual articulation typically used in production of alveolar and retroflex frication in the Singapore variety of Mandarin.

At the same time, it is also worth highlighting that our perception-production studies typically had smaller numbers of male participants as compared to female participants. While this does not affect the overall results of the studies that we have conducted, the lack of a balanced number of male participants in these studies has made it harder for us to investigate gendered differences in speech perception and production of the Singapore Mandarin alveolar-retroflex contrast. Studies on Taiwan Mandarin suggest that there may be gendered extralinguistic factors in retroflex production – such as higher incidences of deretroflexion in Taiwanese men (Rau & Li, 1994; Chuang &

Fon, 2010). On the other hand, early studies on Singapore Mandarin suggest that the opposite may be true, with evidence that Singapore Mandarin-speaking men may instead engage in hypercorrection when producing formal examples of speech (Ng, 1985). Future studies on perception and production of the Singapore Mandarin alveolar-retroflex contrast can focus on recruiting a larger number of male participants to determine if gendered trends can be observed in contemporary Singapore Mandarin speakers.

Next, our study on perception-production links in Chapter 6 showed that there is an apparent link in speech perception and production tied to the retroflex phoneme in Singapore Mandarin. As the sample size for this study was not large, a more representative sample size in future studies may reveal if this link is generalisable to the larger population. At the same time, future follow-up studies on this link could also investigate the effect that perceptual sensitivity training for this contrast has on speech production of the contrast. The results of existing studies suggest that we may find concomitant changes in speech production following perception training (e.g., Sakai & Moorman, 2018). Perceptual training can be conducted with our distributional learning paradigm used in Chapter 7.

8.5 Overall Conclusion

Overall, this project has found evidence in support of the hypothesis that language learning advantages seen in bilinguals are often linked to language-specific transfer effects. This supports other literature showing transfer effects in adults (e.g., Antoniou et al., 2015; Bialystok et al., 2003; Chládková & Šimáčková, 2021) and in children (e.g., Kuo & Anderson, 2012, Kuo et al., 2016). Moreover, our finding that individual differences in Mandarin understanding ability are linked to differences in distributional learning effects further demonstrates that language-specific learning effects can be robust enough to be observed even within the linguistic variations in a single, largely homogenous group of English-Mandarin bilingual adults in Singapore. We have also shown that adults can successfully show distributional learning effects from an unattended training paradigm if there is sufficient overlap between the training stimuli and their known languages. Our detailed investigations into the perception and production of the contemporary Singapore Mandarin alveolar-retroflex contrast also provide fresh insights into the changing phonology of this outer-circle variety of Mandarin. This documentation has valuable implications on different aspects of life in Singapore including in education (e.g., for language acquisition) and healthcare (e.g., for audiometry testing) settings. In line with open science principles, we have made our study stimuli, procedures, data, and analyses available in online repositories for research transparency. This will allow other researchers to conduct further explorations on these topics with the materials that we have created.

References

Anand E., Chung S., & Wagers M. (2011). *Widening the net: Challenges for gathering linguistic data in the digital age.* Research in the Social, Behavioral and Economic Sciences planning activity. https://people.ucsc.edu/%7Eschung/anandchungwagers.pdf

Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A Statistical Basis for Speech Sound Discrimination. *Language and Speech, 46*(2–3), 155–182. https://doi.org/10.1177/00238309030460020601

Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics, 40*, 582–594. http://doi.org/10.1016/j.wocn.2012.05.005

Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. C. M. (2015). The bilingual advantage in phonetic learning. *Bilingualism: Language and Cognition, 18*(4), 683–695. https://doi.org/10.1017/S1366728914000777

Antoniou, M. (2019). The advantages of bilingualism debate. *Annual Review of Linguistics, 5(1)*, 395–415. https://doi.org/10.1146/annurev-linguistics-011718-011820

Atkinson, R. (Executive Producer), & Senior, K. (Executive Producer). (2003). Mr Bean: The animated series. (Series 2). [TV series]. FremantleMedia Enterprises.

Bailey, P. J., & Haggard, M. P. (1980). Perception-Production Relations in the Voicing Contrast for Initial Stops in 3-Year-Olds. *Phonetica, 37*(5–6), 377–396. https://doi.org/10.1159/000260004

Bartolotti, J., Marian, V., Schroeder, S. R., & Shook, A. (2011). Bilingualism and inhibitory control influence statistical learning of novel word forms. *Frontiers in Psychology, 2,* 324. https://doi.org/10.3389/fpsyg.2011.00324

Barrios, S., Rodriguez, J. M., & Barriuso, T. A. (2022). The acquisition of L2 allophonic variants: The role of phonological distribution and lexical cues. *Second Language Research,* [published online]. https://doi.org/10.1177/02676583221099237

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*(3), 305–330. https://doi.org/10.1016/S0095-4470(19)30637-0

Bialystok, E. (2001). Bilingualism in development: Language, literacy, and cognition. Cambridge University Press. https://doi.org/10.1017/CBO9780511605963

Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics, 24*(1), 27–44. https://doi.org/10.1017/S014271640300002X

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. Developmental Science, 7(3), 325–339. https://doi.org/10.1111/j.1467-7687.2004.00351.x

Bialystok, E., & Shapero, D. (2005). Ambiguous benefits: The effect of bilingualism on reversing ambiguous figures. Developmental Science, 8(6), 595–604. https://doi.org/10.1111/j.1467-7687.2005.00451.x

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(3), 859–873. https://doi.org/10.1037/0278-7393.34.4.859

Bialystok, E., & Craik, F. I. M. (2010). Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science, 19*(1), 19–23. https://doi.org/10.1177/0963721409358571

Boersma, P. & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1.31. http://www.praat.org/

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America, 101*(4), 2299–2310. https://doi.org/10.1121%2F1.418276

Broadbent, D. E. (1958). Perception and communication. London: Pergamon Press.

Brysbalrt, M. (1998). Word recognition in bilinguals: Evidence against the existence of two separate lexicons. *Psychologica Belgica, 38*(3–4), 163. https://doi.org/10.5334/pb.932

Callan, D., Callan, A., & Jones, J. A. (2014). Speech motor brain regions are differentially recruited during perception of native and foreign-accented phonemes for first and second language listeners. *Frontiers in Neuroscience, 8*. https://doi.org/10.3389/fnins.2014.00275

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science, 11*(2), 282–298. https://doi.org/10.1111/j.1467-7687.2008.00675.x

Cavallaro, F., & Ng, B. C. (2014). Language in Singapore: From multilingualism to English plus. In J. Hajek & Y. Slaughter (Eds.), *Challenging the Monolingual Mindset* (pp. 33–48). Multilingual Matters. https://doi.org/10.21832/9781783092529-005

Chang, Y. H. (2012). *Variability in cross-dialectal production and perception of contrasting phonemes: The case of the alveolar-retroflex contrast in Beijing and Taiwan Mandarin.* (Doctoral dissertation, University of Illinois at Urbana-Champaign, Illinois, United States). https://core.ac.uk/download/pdf/1020 8504.pdf

Chang, Y.-H. S., Shih, C., & Allen, J. B. (2013). Dialectal variation in the perception of phonological contrasts. *Proceedings of the International Conference on Phonetics of the Languages in China 2013.*

Chang, Y.-H. S., & Shih, C. (2015). Place contrast enhancement: The case of the alveolar and retroflex sibilant production in two dialects of Mandarin. *Journal of Phonetics, 50,* 52–66. https://doi.org/10.1016/j.wocn.2015.02.001

Chen, N. F., Tong, R., Wee, D., Lee, P., Ma, B., & Li, H. (2016) SingaKids-Mandarin: Speech Corpus of Singaporean Children Speaking Mandarin Chinese. *Proceedings of Interspeech 2016*, 1545-1549. http://dx.doi.org/10.21437/Interspeech.2016-139

Chew, P. G. L. (2013). A sociolinguistic history of early identities in Singapore: from colonialism to nationalism. New York: Palgrave Macmillan. https://doi.org/10.1057/9781137012340

Chiu, C., Wei, P.-C., Noguchi, M., & Yamane, N. (2020). Sibilant Fricative Merging in Taiwan Mandarin: An Investigation of Tongue Postures using Ultrasound Imaging. *Language and Speech, 63*(4), 877-897. https://doi.org/10.1177/0023830919896386

Chládková, K., Boersma, P., & Escudero, P. (2022). Unattended distributional training can shift phoneme boundaries. *Bilingualism: Language and Cognition, 25*(5), 827–840. https://doi.org/10.1017/S1366728922000086

Chládková, K., & Šimáčková, Š. (2021). Distributional learning of speech sounds: An exploratory study into the effects of prior language experience. *Language Learning, 71*(1), 131–161. https://doi.org/10.1111/lang.12432

Chong, R. H.-H., & Tan, Y.-Y. (2013). Attitudes toward accents of Mandarin in Singapore. *Chinese Language and Discourse. An International and Interdisciplinary Journal, 4(*1), 120–140. https://doi.org/10.1075/cld.4.1.04cho

Chuang, Y. Y., & Fon, J. (2010). The effect of prosodic prominence on the realisations of voiceless dental and retroflex sibilants in Taiwan Mandarin spontaneous speech. *Proceedings of the 5th International Conference on Speech Prosody,* 100414. https://188.166.204.102/archive/sp2010/papers/sp10_414.pdf

Chung, K. S. (2006). Hypercorrection in Taiwan Mandarin. *Journal of Asian Pacific Communication, 16*(2), 197–214. https://doi.org/10.1075/japc.16.2.04chu

Crosby, D., & Dalola, A. (2021). Phonetic variation in the Korean liquid phoneme. *Proceedings of the Linguistic Society of America 6*(1), 701–712. https://doi.org/10.3765/plsa.v6i1.5002.

Curdt-Christiansen, X. L., & Sun, B. (2016). Nurturing bilingual learners: Challenges and concerns in Singapore. *International Journal of Bilingual Education and Bilingualism, 19*(6), 689–705. https://doi.org/10.1080/13670050.2016.1181606

Department of Statistics, Ministry of Trade and Industry Singapore. (2021). Census of population 2020 statistical release 1: Demographic characteristics, education, language and religion [Datafile]. https://www.singstat.gov.sg/-/media/files/publications/cop2020/sr1/cop2020sr1.pdf

Department of Statistics, Ministry of Trade and Industry Singapore. (2022). Population trends, 2022 [Datafile]. https://www.singstat.gov.sg/-/media/files/publications/population/population2022.ashx

Deterding, D. (2007). Phonetics and Phonology. *Singapore English* (pp. 12–39). Edinburgh: Edinburgh University Press.

de Winter, J. C. F., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research, 44*(2), 147–181. https://doi.org/10.1080/00273170902794206

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*, 781. https://doi.org/10.3389/fpsyg.2014.00781

Dienes, Z., (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology, 72,* 78 – 89. https://doi.org/10.1016/j.jmp.2015.10.003.

Emmorey, K., Borinstein, H. B., Thompson, R., & Gollan, T. H. (2008). Bimodal bilingualism. *Bilingualism: Language and Cognition, 11*(1), 43–61. https://doi.org/10.1017/S1366728907003203

Enomoto, K. (1994). L2 perceptual acquisition: The effect of multilingual linguistic experience on the perception of a ‘less novel’ contrast. *Edinburgh Working Papers in Applied Linguistics, 5,* 15–29. https://eric.ed.gov/?id=ED373553

Escudero, P., Hillenbrand, J., Benders, T., Wanrooij, K., & Hillenbrand, J. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America, 130*(4), EL206–EL212. https://doi.org/10.1121/1.3629144

Escudero, P., & Williams, D. (2014). Distributional learning has immediate and long-lasting effects. *Cognition, 133*(2), 408–413. https://doi.org/10.1016/j.cognition.2014.07.002

Fox, R. A. (1982). Individual variation in the perception of vowels: Implications for a perception-production link. *Phonetica, 39*(1), 1–22. http://doi.org/ 10.1159/000261647

Francis, N. (2016). Language and dialect in China. Chinese Language and Discourse. *An International and Interdisciplinary Journal, 7*(1), 136–149. https://doi.org/10.1075/CLD.7.1.05FRA

Frieda, E. M., Walley, A. C., Flege, J. E., & Sloane, M. E. (2000). Adults’ perception and production of the English vowel /i/. *Journal of Speech, Language, and Hearing Research, 43,* 129-143. http://doi.org/1092-4388/00/4301-0129

Goh, H. L., Styles, S. J., & Onnis, L. (2020, September 9). Effect of unattended distributional training on phoneme category discrimination in English-Mandarin bilingual adult participants in Singapore. https://doi.org/10.17605/OSF.IO/S6VDN

Goh, H. L., Woon, F. T., & Styles, S. J. (2022). Spring Village Corpus of minimal pairs in Mandarin Chinese - Singaporean Adults, V1. [Data]. https://doi.org/10.21979/N9/ZTMPML, DR-NTU

Golestani, N., Paus, T., & Zatorre, R. J. (2002). Anatomical correlates of learning novel speech sounds. *Neuron, 35*(5), 997–1010. https://doi.org/10.1016/S0896-6273(02)00862-0

Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: Reallocation of neural substrates. *NeuroImage, 21*(2), 494–506. https://doi.org/10.1016/j.neuroimage.2003.09.071

Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language, 109*(2–3), 55–67. https://doi.org/10.1016/j.bandl.2008.01.005

Gómez, R. L. (2017). Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1711), 20160054. https://doi.org/10.1098/rstb.2016.0054

Graf Estes, K., Edwards, J., & Saffran, J. R. (2011). Phonotactic Constraints on Infant Word Learning. *Infancy, 16*(2), 180–197. https://doi.org/10.1111/j.1532-7078.2010.00046.x

Gupta, A. F. (1998) The situation of English in Singapore. In J. A. Foley, T. Kandiah, L. Wee, B. Zhiming & A. Fraster-Gupta (Eds.), *English in New Cultural Contexts: Reflections from Singapore* (pp. 106-126). Oxford: Oxford University Press.

Hammarberg, B. (2014). Problems in defining the concepts of L1, L2 and L3. In A. Otwinowska & G. De Angelis (Eds.), *Teaching and Learning in Multilingual Contexts: Sociolinguistic and Educational Perspectives* (pp. 3–18). Multilingual Matters.

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, *23*(1), 65–94. http://doi.org/10.1177/0267658307071601

Hearnshaw, S., Baker, E., & Munro, N. (2018). The speech perception skills of children with and without speech sound disorder. *Journal of Communication Disorders*, *71*, 61–71. https://doi.org/10.1016/j.jcomdis.2017.12.004

Heeren, W. F. L. (2015). Coding pitch differences in voiceless fricatives: Whispered relative to normal speech. *The Journal of the Acoustical Society of America*, *138*(6), 3427–3438. https://doi.org/10.1121/1.4936859

Holland, S. (2021). Data Analysis in the Geosciences: Principal Components Analysis. http://strata.uga.edu/8370/lecturenotes/principalComponents.html

Inceoglu, S. (2016). Effects of perceptual training on second language vowel perception and production. *Applied Psycholinguistics, 37*(5), 1175–1199. https://doi.org/10.1017/S0142716415000533

Jaeger, T. F., & Norcliffe, E. J. (2009). The cross-linguistic study of sentence production. *Language and Linguistics Compass, 3*(4), 866–887. https://doi.org/10.1111/j.1749-818X.2009.00147.x

Jeng, J.-Y. (2006). The acoustic spectral characteristics of retroflexed fricatives and affricates in Taiwan Mandarin. *Journal of Humanistic Studies, 40*(1), 27–48.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: phonetic targets are hyperarticulated. *Language, 69*(3), 505–528. https://doi.org/10.2307/416697

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America, 108*(3 Pt1), 125–1263. https://doi.org/10.1121/1.1288413.

Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics, 21*, 3–28. http://doi.org/10.1016/S0095-4470(19)31319-1

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review, 16*(4), 705–710. https://doi.org/10.3758/PBR.16.4.705

Kawahara, H., Morise, M. (2011) Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework. *Sadhana, 36,* 713–727. https://doi.org/10.1007/s12046-011-0043-3

Ke, H., Pan, L., O'Brien, B. A., & Styles, S. J. (2021). Categorical perception of VOT in multilingual children in Singapore – A preregistered investigation of the CROWN game. *PsyArXiv Preprints*. https://doi.org/10.31234/osf.io/3cgjh

Kemp, C. (2007). Strategic processing in grammar learning: Do multilinguals use more strategies? *International Journal of Multilingualism, 4*(4), 241–261. https://doi.org/10.2167/ijm099.0

Kendall, T., & Fridland, V. (2021). 2 - Sociophonetics and its methods. *In Sociophonetics* (pp. 30–32). Cambridge University Press. https://doi.org/10.1017/9781316809709.003

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language, 42*(6), 703-735. https://doi.org/10.1177/01427237211066405

Kopečková, R. (2016). The bilingual advantage in L3 learning: a developmental study of rhotic sounds. *International Journal of Multilingualism, 13*:4, 410–425. http://dx.doi.org/10.1080/14790718.2016.1217605

Kovács, Á. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences, 106*(16), 6556–6560. https://doi.org/10.1073/pnas.0811323106

Kubler, C. C. (1985). The influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics*, *27*(2), 156–176. https://www.jstor.org/stable/30028064

Kuhl, P. K., Ramírez, R. R., Bosseler, A., Lin, J.-F. L., & Imada, T. (2014). Infants' brain responses to speech suggest Analysis by Synthesis. *Proceedings of the National Academy of Sciences, 111*(31), 11238–11245. https://doi.org/10.1073/pnas.1410963111

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*(4577), 1138-1141. https://doi.org/10.1126/science.7146899.

Kuo, L.-J., & Anderson, R. C. (2008). Conceptual and methodological issues in comparing metalinguistic awareness across languages. In K. Koda & A. Zehler (Eds.), *Learning to read across languages* (pp. 39–67). New York, NY: Routledge.

Kuo, L.-J., & Anderson, R. C. (2012). Effects of early bilingualism on learning phonological regularities in a new language. *Journal of Experimental Child Psychology, 111*(3), 455–467. https://doi.org/10.1016/j.jecp.2011.08.013

Kuo, L.-J., Uchikoshi, Y., Kim, T.-J., & Yang, X. (2016). Bilingualism and phonological awareness: Re-examining theories of cross-language transfer and structural sensitivity. *Contemporary Educational Psychology, 46,* 1–9. https://doi.org/10.1016/j.cedpsych.2016.03.002

Ladefoged, P., & Wu, Z. (1984). Places of articulation: An investigation of Pekingese fricatives and affricates. *Journal of Phonetics, 12*(3), 267–278. https://doi.org/10.1016/S0095-4470(19)30883-6

Ladefoged, P., & Maddieson, I. (1996). The sounds of the world's languages. Oxford, UK: Blackwell Publishers.

Lakens, D., McLatchie, N., Isager, P. M., Scheel, A. M., & Dienes, Z. (2018). Improving inferences about null effects with Bayes factors and equivalence tests. *The Journals of Gerontology Series B Psychological Sciences and Social Sciences, 75*(1), 45 – 57. https://doi.org/10.1093/geronb/gby065.

Lee, C. L., & Phua, C. P. (2020). Singapore bilingual education: One policy, many interpretations. *Journal of Asian Pacific Communication, 30*(1–2), 90–114. https://doi.org/10.1075/japc.00046.lee

Li, F. (2008). *The phonetic development of voiceless sibilant fricatives in English, Japanese, and Mandarin Chinese.* (Doctoral dissertation, The Ohio State University, Ohio, United States). http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.591.4060&rep=rep1&type=pdf

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar.* Berkeley: University of California Press.

Lieberman, A. M., & Mattingly, I, G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1-36. https://doi.org/10.1016/0010-0277(85)90021-6.

Lim, J. J., Chen. S. C., & Hiramoto, M. (2021). “You don’t ask me to speak Mandarin, okay?”: Ideologies of language and race among Chinese Singaporeans. *Language & Communication, 76,* 100–110. https://doi.org/10.1016/j.langcom.2020.10.003

Lin, J. & Moisik, S. (2019). The lingual voice quality settings of Standard Singapore English and Singapore Colloquial English. *Proceedings of the 19th International Congress of Phonetic Sciences,* 206. https://assta.org/proceedings/ICPhS2019/papers/ICPhS_206.pdf

Linares, D., & López-Moliner, J. (2016). quickpsy: An R Package to Fit Psychometric Functions for Multiple Groups. *The R Journal, 8*(1), 122–131. https://journal.rproject.orgarchive/2016-1/linares-na.pdf.

Liu, L., & Kager, R. (2017). Statistical learning of speech sounds is most robust during the period of perceptual attunement. *Journal of Experimental Child Psychology, 164,* 192–208. https://doi.org/10.1016/j.jecp.2017.05.013

Liu, L., Yuan, C., Ong, J. H., Tuninetti, A., Antoniou, M., Cutler, A., & Escudero, P. (2022). Learning to perceive non-native tones via distributional training: Effects of task and acoustic cue weighting. *Brain Sciences,* 12. https://doi.org/10.3390/brainsci12050559

Lock, G. (1988). *Variations, norms and prescribed standard in the Mandarin Chinese spoken in Singapore.* (Doctoral dissertation, University of Sydney, New South Wales, Australia)

Lock, G. (1989). Aspects of variation and change in the mandarin Chinese spoken in Singapore. *Australian Journal of Linguistics, 9*(2), 277-294. https://doi.org/10.1080/07268608908599423

Löfqvist, A., & Mandersson, B. (1987). Long-time average spectrum of speech and voice analysis. *Folia Phoniatrica et Logopaedica, 39*(5), 221–229. https://doi.org/10.1159/000265863

Luo, S. (2020). Articulatory tongue shape analysis of Mandarin alveolar–retroflex contrast. *The Journal of the Acoustical Society of America, 148*(4), 1961–1977. https://doi.org/10.1121/10.0002111

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101–B111. https://doi.org/10.1016/S0010-0277(01)00157-3

Mathôt, S., Schreij, D., Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324.

McAuliffe, M. (2017, February 23). How to create synthetic speech continua using STRAIGHT. https://memcauliffe.com/how-to-create-synthetic-speech-continua-using-straight.html

McGowan, R. S., Nittrouer, S., & Manning, C. J. (2004). Development of [ɹ] in young, Midwestern, American children. *The Journal of the Acoustical Society of America, 115*(2), 871–884. https://doi.org/10.1121%2F1.1642624

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination: Acoustic salience and developmental speech perception. *Developmental Science, 13*(3), 407–420. https://doi.org/10.1111/j.1467-7687.2009.00898.x

National Library Board, Government of Singapore. (2016, August 31). Bilingual policy. https://eresources.nlb.gov.sg/infopedia/articles/SIP_2016-09-01_093402.html

Newman, J. (1982). A study of Hokkien-Mandarin phonological correspondences: Implications for the teaching of Mandarin to Hokkien speakers. *Occasional Papers (Regional Language Centre)* (No. 22). Singapore: SEAMEO Regional Language Centre.

Newman, J. (1983). Hokkien-Mandarin Phonological Correspondences as Potential Transfer Strategies. In Eppert, F. (Ed.), *Transfer and Translation in Learning and Teaching* (pp. 90–103). Singapore: Singapore University Press.

Newman, J. (1988). Singapore’s speak Mandarin Campaign. *Journal of Multilingual and Multicultural Development, 9*(5), 437–448. https://doi.org/10.1080/01434632.1988.9994348

Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *The Journal of the Acoustical Society of America, 113*(5), 2850–2860. https://doi.org/10.1121/1.1567280

Ng, P. C. L. (2014). A study of attitudes towards the speak Mandarin Campaign in Singapore. *Intercultural Communication Studies, 23*:3 (pp. 53–65. Singapore: Springer

Ng, B. C. (1985). A Study of the Variable /sh/ in Singapore Mandarin. In Bradley, D. (Ed.) *Papers in Southeast Asian Linguistics No. 9: Language Policy, Language Planning, and Sociolinguistics in South-east Asia* (pp. 31–37). Canberra: Pacific Linguistics

Oh, J. S., Au, T. K.-F., & Jun, S.-A. (2010). Early childhood language memory in the speech perception of international adoptees. *Journal of Child Language, 37*(5), 1123–1132. https://doi.org/10.1017/S0305000909990286

Ou, S., & Guo, Z. (2014). Mandarin retroflex sounds perceived by non-native speakers*. Chinese Collections of National Taipei University of Education,* 42 – 76. https://tpl.ncl.edu.tw/NclService/pdfdownload?filePath=lV8OirTfsslWcCxIpLb

UfvnJVVyS2MdWoteG0es3wArWL458qQ1UxEbw9UWP4zbv&imgType=Bn5sH4BGpJw=&key=9_Jt0kA3BKHdUYtXbhriYY2qqWybcx2_q1C2oxESAeVVU9OyINO4qBZJhLTxWd&xmlId=0006 794641

Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology, 26*(6), 615–639. https://doi.org/10.1080/20445911.2014.944914

Pan, L., Ke, H., & Styles, S. J. (2021). Early linguistic experience shapes bilingual adults' hearing for phonemes in both languages. *PsyArXiv Preprints*. https://doi.org/10.31234/osf.io/qvsu3

Pakir, A. (1993). Two tongue tied: Bilingualism in Singapore. *Journal of Multilingual and Multicultural Development, 14*(1–2), 73–90. https://doi.org/10.1080/01434632.1993.9994521

Pakir, A. (2008). English as a Lingua Franca: Negotiating Singapore's English Language Education. In S. M. Wu, T. R. F. Tupas, M. L. Chew, M. L. C. Sadorra & C. Varaprasad (Eds.), *The English Language Teaching and Learning Landscape: Continuity, Innovation and Diversity* (pp. 21–32). Singapore: Centre for English Language Communication, NUS.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24,* 175–184. https://doi.org/10.1121/1.1906875

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team (2022). nlme: Linear and nonlinear mixed effects models. R package version 3.1-155. https://CRAN.R-project.org/package=nlme.

Platt, J. (1980). Multilingualism, polyglossia, and code selection in Singapore. In E. A. Afendras & E. C. Y. Kuo (Eds.), *Language and Society in Singapore* (pp. 63-86). Singapore: Singapore University Press.

Punišić, S., Bilibajkić, R., Vojnović, M., & Subotić, M. (2019). Perception of speech sounds’ pronunciation quality based on the acoustic features. *Proceedings of the 7th International Conference on Fundamental and Applied Aspects of Speech and Language.* 305–313. http://www.iefpg.org.rs/Conference/2019/S&L2019_PROCEEDINGS.pdf

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rau, D.-H., and M.-C. Li. (1994). Phonological Variation of /tʂ/, /tʂh/, /ʂ/ in Mandarin Chinese. *Proceedings of the Fourth International Conference on Chinese Language Pedagogy.* 346–366.

Reh, R. K., Hensch, T. K., & Werker, J. F. (2021). Distributional learning of speech sound categories is gated by sensitive periods. *Cognition, 213,* 104653. https://doi.org/10.1016/j.cognition.2021.104653

Rvachew, S., & Jamieson, D. G. (1989). Perception of Voiceless Fricatives by Children with a Functional Articulation Disorder. *Journal of Speech and Hearing Disorders, 54*(2), 193–208. https://doi.org/10.1044/jshd.5402.193

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294):1926–1928. http://doi.org/10.1126/science.274.5294.1926.

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics, 39(*1), 187–224. https://doi.org/10.1017/S0142716417000418

Sanz, C. (2000). Bilingual education enhances third language acquisition: Evidence from Catalonia. Applied Psycholinguistics, 21(1), 23–44. https://doi.org/10.1017/S0142716400001028

Schmalz, X., Biurrun Manresa, J., & Zhang, L. (2023). What is a Bayes factor? *Psychological Methods, 28*(3), 705–718. https://doi.org/10.1037/met0000421

Shih, Y., & Kong, E. (2011). Perception of Mandarin fricatives by native speakers of Taiwan Mandarin and Taiwanese. *Proceedings of the 23rd North American Conference on Chinese Linguistics*, *1*, 110-119. https://naccl.osu.edu/sites/naccl.osu.edu/files/NACCL-23_1_08.pdf

Singh, L., Fu, C. S. L., Tay, Z. W., & Golinkoff, R. M. (2018). Novel word Learning in bilingual and monolingual infants: Evidence for a bilingual advantage. *Child Development, 89*(3), e183–e198. https://doi.org/10.1111/cdev.12747

Singh, L., & Seet, S. K. (2019). The impact of foreign language caregiving on native language acquisition. *Journal of Experimental Child Psychology, 185*, 51–70. https://doi.org/10.1016/j.jecp.2019.04.010

Soh, K. W., & Loo, J. H. Y. (2020). A review of Mandarin speech recognition test materials for use in Singapore. *International Journal of Audiology, 60*(6), 399–411. https://doi.org/10.1080/14992027.2020.1826587

Soh, K. W., & Loo, J. H. Y. (2023). Development of phonologically-balanced and perceptually equivalent Singapore Mandarin word lists for word recognition test. *Proceedings of Singapore Healthcare, 32*, 201010582311784. https://doi.org/10.1177/20101058231178402

Sóskuthy, M. (2017). Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction. *ArXiv, 1703.05339*. https://doi.org/10.48550/arXiv.1703.05339

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*(3), 281-284. https://doi.org/10.1111/1467-9280.00151

Starr, R. L. (2022). Production and evaluation of sociolinguistic variation in Mandarin Chinese among children in Singapore. In R. Bayley, D. R., Preston,

& X, Li (Eds.), *Variation in Second and Heritage Languages* (pp. 43–70). Amsterdam: John Benjamins.

Starr, R. L., & Hiramoto, M. (2019). Inclusion, exclusion, and racial identity in Singapore's language education system. *International Journal of Applied Linguistics, 29*(3), 341–355. https://doi.org/10.1111/ijal.12242

Starr, R. L., & Kapoor, S. (2021). "Our graduates will have the edge": Linguistic entrepreneurship and the discourse of Mandarin enrichment centers in Singapore. *Multilingua, 40*(2), 155–174. https://doi.org/10.1515/multi-2020-0033

Starr, R. L. & Wang, T. (2021). Navigating L2 sociolinguistic variation amid contested norms and societal shifts: A case study of two L2 Mandarin speakers in Singapore. In A. Nardy, A. Ghimenton, & J-P. Chevrot (Eds.) *Sociolinguistic variation and language across the lifespan* (pp. 199–226). Amsterdam: John Benjamins.

Stevens, K. N., & Halle, M. (1967). Remarks on analysis by synthesis and distinctive features. In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form* (pp. 88–102). Cambridge, MA: MIT Press.

Stuart-Smith, J. (2020). Changing perspectives on /s/ and gender over time in Glasgow. *Linguistics Vanguard, 6*(s1), 20180064. https://doi.org/10.1515/lingvan-2018-0064

Sze, W. P., Rickard Liow, S. J., & Yap, M. J. (2014). The Chinese Lexicon Project: A repository of lexical decision behavioral responses for 2,500 Chinese characters. *Behavior Research Methods, 46*(1), 263–273. https://doi.org/10.3758/s13428-013-0355-9

Teo, P. (2005). Mandarinising Singapore: A critical analysis of slogans in Singapore’s ‘Speak Mandarin’ campaign. *Critical Discourse Studies, 2*(2), 121–142. http://doi.org/10.1080/17405900500283565

Terry, J., Ong, J. H., & Escudero, P. (2015). Passive distributional learning of non-native vowel contrasts does not work for all listeners. *Proceedings of the 18th International Congress of Phonetic Sciences.* https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0867.pdf

Tiede, M. K., Chen, W.-R., & Whalen, D. H. (2019). Taiwanese Mandarin sibilant contrasts investigated using coregistered EMA and ultrasound. *Proceedings of the 19th International Congress of Phonetic Sciences.* https://icphs2019.org/icphs2019-fullpapers/pdf/full-paper_690.pdf

Thompson, A. S. (2015). Are Your Participants Multilingual? The role of self-assessment in SLA research. *Language in Focus, 1*(1), 51–65. https://doi.org/10.1515/lifijsal-2015-0004

Tremblay, M.-C., & Sabourin, L. (2012). Comparing behavioral discrimination and learning abilities in monolinguals, bilinguals and multilinguals. *The*

Journal of the Acoustical Society of America, *132*(5), 3465–3474. https://doi.org/10.1121/1.4756955

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development, 75*(4), 1067–1084. https://doi.org/10.1111/j.1467-8624.2004.00726.x

Tse, C.-S., Yap, M. J., Chan, Y.-L., Sze, W. P., Shaoul, C., & Lin, D. (2017). The Chinese Lexicon Project: A megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words. *Behavior Research Methods, 49*(4), 1503–1519. https://doi.org/10.3758/s13428-016-0810-5

Tso, R. P. R. (2017). *The effect of Chinese characters on the speech perception and production of retroflex sibilants in Taiwanese Mandarin.* (Doctoral dissertation, Rice University, Texas, United States). https://scholarship.rice.edu/bitstream/handle/1911/96177/TSO-DOCUMENT 2017.pdf

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., & Best, C. (1994). Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6-12 months: Language-specific developmental changes in speech perception abilities. *3rd International Conference on Spoken Language Processing (ICSLP 1994)*, 1695–1698. https://doi.org/10.21437/ICSLP.1994-438

Vandermosten, M., Wouters, J., Ghesquière, P., & Golestani, N. (2018). Statistical learning of speech sounds in dyslexic and typical reading children. *Scientific Studies of Reading, 23*(1), 116–127. https://doi.org/10.1080/10888438.2018.1473404

van Dommelen, W. A. (2019). Is the voiceless palatal fricative disappearing from spoken Norwegian? *Proceedings of the 19th International Congress of Phonetic Sciences*. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_814.pdf

van Rij, J., Wieling, M., Baayen, R. H, & van Rijn, H. (2020). itsadug: Interpreting time series and autocorrelated data using GAMMs. R package version 2.4. https://cran.r-project.org/web/packages/itsadug/itsadug.pdf

Verma, R, & Chawla, P. (2003). Comparative analysis of Hindi retroflex and dental CV syllables and their synthesis. *Workshop on Spoken Language Processing An ISCA-Supported Event*, Mumbai, India. https://www.isca-speech.org/archive_open/wslp_03/wslp_163.pdf

Wang, T., & Saffran, J. R. (2014). Statistical learning of a tonal language: The influence of bilingualism and previous linguistic experience. *Frontiers in Psychology, 5,* 953. https://doi.org/10.3389/fpsyg.2014.00953

Wanrooij, K., Escudero, P., & Raijmakers, M. E. J. (2013). What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics*, *41*(5), 307–319. https://doi.org/10.1016/j.wocn.2013.03.005

Wanrooij, K., Boersma, P., & van Zuijen, T. L. (2014). Fast phonetic learning occurs already in 2-to-3-month old infants: An ERP study. *Frontiers in Psychology, 5,* 77. https://doi.org/10.3389/fpsyg.2014.00077Wells, J. (1982). *Accents of English*. Cambridge: Cambridge University Press.

Werker, J. F. (1986). The effect of multilingualism on phonetic perceptual flexibility. *Applied Psycholinguistics, 7*(2), 141–155. https://doi.org/10.1017/S0142716400007360

Werker, J. F., & Tees, C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development,7,* 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How Do Infants Become Experts at Native-Speech Perception? *Current Directions in Psychological Science, 21*(4), 221–226. https://doi.org/10.1177/0963721412449459

Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Applied Psycholinguistics, 39*(4), 703–728. https://doi.org/10.1017/S0142716418000152

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modelling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. Journal of Phonetics, 70, 86–116. https://doi.org/10.1016/j.wocn.2018.03.002

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B), 73*(1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Wu, C-Y., O’Brien, B. A., Styles, S. J., & Chen, S-H. A. (2020). The impact of bilingualism on skills development and education. In S. C. Tan & SH. A. Chen (Eds.), *Transforming Teaching and Learning in Higher Education* (pp. 47–69). Singapore, Singapore: National Institute of Education.

Xie, W., & Cavallaro, F. (2016). Attitudes towards Mandarin–English bilingualism: A study of Chinese youths in Singapore. *Journal of Multilingual and Multicultural Development, 37*(6), 628–641. https://doi.org/10.1080/01434632.2015.1122603

Yang, E. (2017). Bilinguals’ working memory (WM) Advantage and their dual language practices. *Brain Sciences, 7*(7), 86. https://doi.org/10.3390/brainsci7070086

Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy, 15*(4), 420–433. https://doi.org/10.1111/j.1532-7078.2009.00024.x

Yoshida, H., Tran, D. N., Benitez, V., & Kuwabara, M. (2011). Inhibition and adjective learning in bilingual and monolingual Children. *Frontiers in Psychology, 2,* 210. https://doi.org/10.3389/fpsyg.2011.00210

Zlatin, M. A., & Koenigsknecht, R. A. (1976). Development of the Voicing Contrast: A Comparison of Voice Onset Time in Stop Perception and Production. *Journal of Speech and Hearing Research, 19*(1), 93–111. https://doi.org/10.1044/jshr.1901.93

Appendices

Appendix 3.1

Table A1 *List of English words for follow-up ultrasound study*

No.	Word
1	Cat-seat
2	Catsit
3	Catsuit
4	Cheap
5	Cheat
6	Chin
7	Cho
8	Choo
9	Chop
10	Seat
11	Seep
12	Sheep
13	Sheet
14	Shin
15	Shoe
16	Shop
17	Show
18	Shun

19	Shy
20	Sin
21	Sigh
22	So
23	Soon
24	Sun
25	Sop

Appendix 4.1

Growing Collection Consent Form

The Growing Collection of Human Voices Release Form for Vocal Recordings of Speech

Thank you for allowing us to record your voice for the current study. We plan to use the recording in the manner described to you previously in the Information Sheet: Only the investigators involved in the project will hear your vocal recording, your identity will be kept anonymous, and the digital file will be deleted after the conclusion of the study. However, interested participants might like to consider releasing their digital files into an ongoing collection designed to maximize the impact of your contribution.

The Growing Collection of Human Voices

It is one of our goals to build up a collection of vocal recordings from different communities around the world, which will help us to answer questions about the function and features of speech produced for different listeners. We are particularly interested in how child directed speech might highlight aspects of language which might facilitate learning, and how these features might differ from other types of speech. The more vocal recordings we add to this collection, the more powerful it will be in helping us to answer these questions. We would like to invite you to join the Growing Collection, and have prepared this document to help you understand what is involved. We will ask you for a set of permissions, and in return, we provide you with a set of assurances. More details can be found on the reverse of this page.

The reason we are asking for special permission is that every human voice is unique, rich, and special. This means that there is a possibility that at some future time, someone may hear your voice and think that they recognize you as the speaker. We therefore cannot guarantee complete anonymity in this collection, but we will do our best to protect you.

Permissions requested:

- Permission to retain your recording and information
- Permission to play recordings to participants in future studies
- Permission to present recordings in public
- Permission to share recordings with other registered researchers

Assurances:

- Your choice of anonymity or identity
- Respect

Permission to retain your recording and information

We would like to add your recorded voice to the Growing Collection of vocal recordings designed to investigate different aspects of Child Directed Speech, and how it compares to speech across different contexts. For this reason, we would like permission to keep the recording of your voice and the other information we collected about you, instead of deleting all files after the end of the study. In this way, when future research questions arise, your vocal sample could help us to answer those questions as well. The primary records of the Growing Collection of Human Voices will be stored on password protected PCs of the Growing Collection manager, currently Asst Prof Suzy Styles, at Nanyang Technological University, Singapore. The manager of the Growing Collection may change in the future, but you can find details on the website https://researchdata.ntu.edu.sg/dataverse/GrowingCollection/

Permission to play recordings to participants in future studies

Since it is difficult for us to predict all possible research possibilities which may arise as the outcome of our current study, we would like permission to play your vocal recordings to participants in future studies. For example, vocal recordings might be used in guessing games (e.g., "guess which picture this recording was describing"); vocal recordings might be combined with story-book pictures, so that the eye-movements of adults or children can be monitored while they listen and watch the story; or samples from the vocal recordings might be evaluated for acoustic properties such as intonation contour, affective intensity, and congruence.

Permission to present recordings in public

When we have concluded the current study, and analyzed our results, it is sometimes useful to accompany a scientific description of the findings with some examples to illustrate the method and results. These kinds of illustrations can be helpful to the scientific community, and also help to show parents what kind of work we do, and what they might expect if their child were to take part in one of our studies. For this reason, we would like permission to present your recording to members of the broader public, as an illustration of our method and/or results. In some contexts, this may mean transferring broadcast rights to a third party (for example a scientific journal, or a documentary film company), so that they may include the recording in their online audio-visual publications, or their broadcasts. In the case of rights transfer, only vocal recordings and non-identifiable information will be transferred. Your information/recordings will not be sold for profit.

Permission to share recordings with other researchers

In the future, the recordings and information in the Growing Collection may have relevance outside the interests of the current investigators. For this reason, we would like your permission to share your recordings with other researchers, so that your recordings can have the greatest possible scientific impact. In order to protect you from risk, embarrassment, or any form of unpleasantness, all current and future investigators who are allowed access to the Growing Collection, will be required to abide by the conditions laid out in the sections below. Please note that this means copies of your recordings may be stored on sites other than those listed above.

Assurance of anonymity or identity

In general, we will not store your real name with any of the information or recordings collected from you. However, since every human voice is unique, there is the possibility that a listener may think they recognize your voice when they hear your recording. We will do our best to keep your identity anonymous, but there is always the potential for re-identification. The recordings you have made contain no sensitive, unpleasant or individually identifying information, so the potential risk associated with re-identification is considered to be quite low. However, you can decide whether you are comfortable with this risk. On the other hand, since the recordings represent your unique vocal style, you might like to know when a sample of your speech has been used as an illustration. For this reason, we allow you to select a Username, which will appear with your voice any time it is used in illustration. For example, you might like your recordings to be known as TomsMum, MiniKate, or a stage name, if you are an actor. Clearly, selecting a username which identifies you to others could make it impossible to ensure your anonymity, but we leave this choice up to you.

Assurance of respect

Any researcher who joins the project or requests access to the recordings will have to agree to the following conditions before being allowed access to the recordings: Researchers must give their name and current research affiliations (school, university, or equivalent research organization). Recordings of voices must be treated with respect, and should not be presented in any context which might cause harm or embarrassment to the speaker. For example recordings should not be associated with assessments of racial prejudice, evaluations of likely criminality, sexual orientation, religious affiliation, or any other sensitive material. Recordings should not be paired with distressing or unpleasant stimuli in another sensory domain (e.g., unpleasant pictures, unpleasant smells). No individual should be identified as ‘bad at’ any aspect of the task. Where Usernames have been given, Usernames must be presented alongside any vocal samples used as illustrations of method or results. For example, named or listed in the credits of a documentary; named in a digital file published as supplementary material in a journal article; or listed in live demonstrations (e.g., Presentation at academic conferences, Public science lectures). Publications and presentations arising from access to the Growing Collection must follow the citation guidelines available on the website: https://researchdata.ntu.edu.sg/dataverse/GrowingCollection/

Your Agreement:

I have received the relevant information and agree to release the recording of my voice, and non-identifiable information collected at the time of the recording, into the Growing Collection of Human Voices.

o I agree (1)

o I disagree (2)

Please enter your Participant ID here:

Appendix 7.1

Language ID Mandarin Background Questions

Please tell us a bit about your Mandarin experience by answering the following questions.

How old were you when you first started **understanding** Mandarin?

____

____

How old were you when you first started **speaking** Mandarin?

____

____

How many years did you formally study Mandarin? (i.e., preschool, school, and university)

____

____

How old were you when you first began formally studying Mandarin?

____

____

When you consume media (e.g., music, TV shows, movies), what percentage of the media is in:

When you communicate with people, what percentage of your conversation is in: