Cross-language Differences in Fricative Processing and Their Influence on Non-native Fricative Categorisation

Yue Zheng

A thesis submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Speech, Hearing and Phonetic Sciences

University College London

2021

Declaration

I, Yue Zheng, declare that the thesis is composed by myself, the work presented in this thesis is my own, and it has never been submitted for any other degree or programme. Where reference has been made to the work of others, I confirm that appropriate credit has been given within this thesis.

Acknowledgements

First and foremost, I sincerely thank my primary supervisor, Bronwen Evans, who has always been so encouraging and inspirational to me. I will be forever grateful for her taking me on as her student when it was almost the end of my PhD. I really could not have finished this thesis without her. She was always there when I needed help, full of good ideas and valuable advice. Her cheering attitude always manages to relieve my stress and anxiety a bit, and pushes me to keep plodding on. She will always be my role model as a successful woman in research. I would also like to thank my secondary supervisor Mark Huckvale, who is always responding, so promptly, patiently, and informatively, to my emails full of naïve questions about statistics and coding. The countless email trains in my inbox are evidence of the expertise and patience of him as a great professor, and I will be forever grateful for his help. I would like to thank my previous supervisor, Paul Iverson, who helped me a lot with getting into this programme with a scholarship. I was fortunate to have learned a lot from him about L2 perception, EEG, and how to read MATLAB and Praat code.

Many thanks go to all the lovely and intelligent researchers and fellow PhD students, both previous and current, working in Chandler House: Jieun Song, for all her help with EEG and coding with MATLAB, and emotional support—I treasure all the coffee breaks and chats we had in and around the office; Katharina Zenke, for being a good friend and for all her company through good and bad times—I feel so fortunate to have gone through them with her by my side (both literally and figuratively); Faith Chiu, for checking on me and thinking of me even after she left UCL; Alice Wu, for her help with R and for all the pleasant chats we had during lockdown; Shego Wu and Anqi Xu, for giving me the strength and inspiration that I needed when I was going through rough patches. I would also like to thank Gwen Brekelmans, Emma Brint, Gisela Tomé Lourido, Gwijde Maegherman, Shiran Koifman (and Maddie!), Clara Liu, Xiao Fu, Han Wang, Julie Saigusa, Luke Martin, Anna Exenberger, Max Paulus, Ana Campos Espinoza, Elise Kanber, and Bryony Payne, for creating a supporting and inspirational working environment.

I am grateful for all the PALS staff for their help and support throughout my PhD, and for that running into them in the building and having a chat with them can always

brighten up a day. Thanks go to Richard Jardine, for his patience and help to solving my big and small problems and requests all these years; and to Andrew Clark, for always being there to make technical problems during testing disappear, and for being a good listener when I expressed my frustration with testing and other stuff. I am also grateful for Dave Cushing, Steve Newton and the IT team, and the Finance team. My gratitude also goes to Saiid Omar, for always greeting me with a smile whenever I arrive at Chandler House, and for taking good care of my orchid during lockdown.

This thesis would not be in this state without the help from all my participants, so special thanks go to them. My thanks also go to Andy Chin, Jeremy Ng, and Karen Li, for helping me recruit Cantonese speakers from Hong Kong.

No words can describe how grateful I am to my family, for their unwavering faith in me, their unconditional support (both emotionally and financially) through all these years. Being apart does not make their love seem any weaker. I could not have gotten this far without them.

I am more than grateful for Adam Williams, for him always being at the receiving end of my frustration, grumpiness, and other negative emotions without (too many) complaints these years. I can always rely on him for a boost of confidence and a little push when I am too tired to carry on. Thank you for having faith in me, and for all your effort to make my life easier and happier.

My thanks also go to my amazing viva examiners: Chris Carignan and Ocke Bohn. They have given me valuable advice, kind encouragement, and altogether an enjoyable viva experience.

I would like to save the last paragraph to myself, to whom I should also be grateful. A pat on my back for never giving up, and for growing stronger and wiser after all the difficulties I experienced.

Abstract

Studies have shown that native speakers of Mandarin Chinese and Hong Kong Cantonese tend to have difficulty perceiving the English fricative /θ/. However, although both languages have /f/ and /s/ categories, Mandarin speakers tend to assimilate it to their /s/ category whilst Cantonese speakers would assimilate it to their /f/ category. Over three studies, this thesis investigated various factors that may lead to this difference, while enhancing our understanding of the acoustics and the perception of the fricatives of these languages.

Study 1 explored acoustic properties of target fricatives of the three languages (Mandarin, Cantonese, English) using audio recordings from native speakers, and conducted comparisons of the fricatives within and across languages. The results showed that the phonemes /f s/, even though shared by the three languages, were produced differently in the different languages, likely due to the effects of the different fricative inventories. Moreover, different acoustic cues were more or less effective in distinguishing between the different fricatives in each language, indicating that native speakers of these languages likely rely on these cues differently. Study 2 examined how transition cues may affect the identification of /f/ and /s/ by native speakers of the respective languages by combining a phoneme monitoring task and EEG measures. Target fricatives were spliced with vowels to create stimuli with congruent or incongruent transitions. In contrast to previous studies (e.g., Wagner, Ernestus & Cutler, 2006), the results revealed that all groups attended to formant transitions when processing fricatives, despite their differing native fricative inventory sizes. Study 3 investigated cross-language differences in categorisation boundaries of target fricative pairs using a behavioural identification task. The study interpolated pairs of stimuli to create a frication continuum and a vowel continuum, forming a 2-dimensional stimuli grid. The results indicated that frication was the primary cue for fricative identification for the native English, Cantonese, and Mandarin speakers, but also revealed cross-language differences in fricative boundaries. Overall, the results of these studies demonstrate that the processing of fricatives was largely driven by the frication section, and the differential assimilation of /θ/ was likely due to the different acoustics of the same fricative category across languages. The results also motivate a reconsideration of the role of coarticulatory cues in fricative perception.

Impact Statement

The research presented in this thesis enhances our understanding of native and non-native speech sound processing. The results could be beneficial not only to future research in disciplines including speech recognition and comprehension, but also to updating theories and methodologies in second language teaching/learning, as well as automatic speech recognition. In addition, the results could also have a positive impact on social inclusion and inter-cultural communication.

This project has shown that speech processing strategies are not only language-specific, but also involve more than just acoustic factors in a language's phonological system than previous studies have argued. This discovery is a challenge to cross-linguistic studies in speech recognition, as it reveals the importance to take on a more holistic view of L2 sound perception, rather than just analysing at the phonetic level. More importantly, it enhances awareness of the complexity of language comprehension: the fact that accurate perception relies on a complex set of phonological factors should be carefully considered when designing future experiments. In addition, the attempt to model speech sound processing using machine-learning techniques also provides insights into how we can improve automatic speech processing and recognition modelling, especially for understanding accented speech.

The new knowledge generated by the project demonstrates how one's language experience affects speech processing, and how this can create difficulties in second language learning. The findings will specifically help address the challenges faced by Chinese (Mandarin and Cantonese) speakers when learning English. Chinese learners of English are one of the largest in the world (estimated to be at least 395 million by the year 2006, Population By-census, 2016; Wei & Su, 2012) and this thesis has made it clear that Chinese speakers process speech in a different way from English speakers, due to the differences in the phonological systems of the languages. The findings will be useful in improving existing second language teaching techniques for Chinese-speaking English learners, especially in improving methods to tackle listening and pronunciation difficulties of English fricative sounds. In this way, this project benefits a very large population through improving their learning experience and efficiency.

The positive social and cultural impact yielded by this project includes enhancing social inclusion in English-speaking countries. The new knowledge facilitates mutual understanding among people from different language backgrounds of the challenges faced when learning a new language. This in turn has the potential to increase awareness of difficulties and to generate greater empathy towards non-English speaking immigrants.

The discoveries of this project have and will be made accessible to a wide range of interested groups in academia, both inside and outside the discipline. To be specific, this thesis has been/will be presented in part or as a whole in a variety of conferences and seminars, including Attention to Sounds 2018, International Congress of Phonetic Sciences 2019, and Women in Research Alumni Conference 2021. The project has received strong interest from experts in language education. Potential collaboration with The Education University of Hong Kong on developing a training technique that improves English sound perception is in the planning stages.

Contents

List of Figures

List of Tables

Chapter 1 General Introduction

During the learning process of a native language, young learners - who once were all universal listeners - quickly develop a system to focus on acoustic cues that are crucial to forming perceptual contrasts for native phonemes. Almost simultaneously, they deem some cues irrelevant for distinguishing native sounds, and ignore them in speech. Difficulties occur when the acoustic cues they have learned to neglect happen to be crucial to phonemic contrasts of a foreign language they start to study. In the field of research on second language (L2) speech sound perception, studies have established a view which believes native language (L1) experience influences L2 sound processing. In fact, one of the main challenges that L2 learners face when perceiving L2 sounds is posed by perceptual specialisation for L1 phonemes which facilitates L1 processing (Cutler, 2000; Flege, 2002). The processing of speech sounds relies on acoustic cues such as burst noise, voice onset time, frication noise, formant information, and coarticulatory cues, and the native language knowledge modifies the processing strategy towards the cues that are important for signalling a difference between phonemes in the native language. A native listening strategy may either facilitate or inhibit accurate perception and categorisation of L2 sounds. When L1 experience leads to inaccurate categorisation of L2 sounds, it may reduce the intelligibility of one's L2 speech, and may cause communication difficulties. It is thus crucial to discover the details of how L1 knowledge and listening strategy interact with various types of L2 sounds.

Investigating L1 listening strategies and their interaction with L2 input was primarily focussed on providing answers for specific L2 learning difficulties. The differences and similarities between the L1 and L2 phonemic inventories are usually emphasised on, and the predictions of assimilation are made at a phonemic level. Models have been proposed to describe the interaction between L1 and L2 phonemic systems. Two widely discussed models are the Speech Learning Model (SLM) and Perceptual Assimilation Model (PAM), proposed by Flege (1995) and Best (1995) respectively. While PAM was originally set out to explain L2 sound assimilation by naïve listeners, PAM-L2 should also be included in the discussion of L2 listening difficulties (Best & Tyler, 2007). SLM proposes that L1 and L2 phonetic elements exist in a common phonological space, so when an L2 phoneme is greatly similar acoustically to an L1

counterpart, their sound categories will merge into one (Flege, 1995). According to SLM, the perception of L2 phonetic segments is predictable based on the knowledge of the learner's L1 phonemic system. Similarly, PAM asserts that when an L2 phoneme is similar to an L1 sound, the L2 sound will be assimilated into the L1 segmental category, while the similarity is defined at a gestural level instead of at a phonetic level, such as similarities in terms of the place of articulation, active articulators, constriction degree, and gestural phasing (Best, 1995). For L2 learners, on the other hand, PAM-L2 predicts that the perceptual similarities may be defined at a gestural, phonetic, and/or phonological level (Best & Tyler, 2007).

All models agree that L2 phoneme categorisation largely depends on whether there is a similar category present in L1 phoneme inventories, though the models have disagreements on what is the most important aspect for L2 listeners that determines perceptual similarity. Based on empirical evidence, it appears that such a decision depends on a listener's language experience. Languages have different criteria for making the "which is more similar" judgement, depending on which acoustic cues are more important than others for identifying a target phoneme. As a result, learners from various L1 backgrounds may categorise the same L2 phoneme differently.

English, the most learned L2, can present difficulties for learners with different L1s. Its relatively rich fricative inventory can be especially hard to perceive for those from other language backgrounds. Facing English fricatives, L2 listeners with smaller native fricative inventories tend to assimilate them to fricatives from their L1. However, these assimilation patterns can differ across languages, even ones with similar fricative inventories. For example, Japanese and Russian both lack a dental fricative, and have the same potential phoneme categories (/s t t͡s/) to which English /θ/ could be assimilated. Interestingly, native Japanese speakers tend to assimilate /θ/ to their /s/ category, while native Russian speakers assimilate /θ/ to their /t/ category (Weinberger, 1997). One may argue that Japanese and Russian are distinct languages that do not have a close connection, but the differential assimilation taking place in European and Canadian French for /θ/ appears harder to explain. Both types of French share the same fricative inventory, but native speakers of the respective dialects systematically assimilate /θ/ differently: Canadian French speakers assimilate it to their /t/ category, but European French speakers tend to assimilate it to their /f/ or /s/

category (Picard, 2002; Tyler et al., 2019). These instances indicate that the language-specific assimilation of non-native sounds involves processing of acoustic cues within phonemes, as only observing the differences between inventories at a phonemic level cannot explain any of the variations mentioned above. This phenomenon that different languages would replace the same L2 sound with different phonemes during their L2 production learning is referred to as "differential substitution" by Weinberger (1997) from a production point of view; according to PAM and PAM-L2, this phenomenon may be referred to as differential assimilation from a speech perception point of view.

The focus of this thesis is the differential assimilation that takes place in Mandarin Chinese (later referred to as Mandarin) and Hong Kong Cantonese (later referred to as Cantonese) of the English voiceless dental fricative /θ/. There have been a few studies that have mentioned the difficulty of processing this dental fricative for Mandarin and Cantonese speakers, since it is absent in both languages. Studies have reported that native Mandarin speakers tend to perceive and produce /θ/ as their native /s/ category (Eaves, 2011; Jiang, 1995; Liang, 2014), while native Cantonese speakers treat /θ/ as their /f/ category (Hung, 2000; Meng, Lo, Wang & Lau, 2007; Meng, Zee, Lee & Lee, 2007; Peng & Setter, 2000). A question arises when comparing the fricative inventories of Mandarin and Cantonese: as shown in Table 1-1, the two languages both have /f/ and /s/, so what motivates them to choose differently? Some studies have noted this difference (Deterding, 2006; Hung, 2000); nevertheless, studies had yet provided a detailed explanation accompanying experimental analyses to address this question. This thesis attempts to bridge the gap between the observation and the reasons behind the differential substitution, by addressing the research question stated below:

> Why do Mandarin and Cantonese speakers perceive the English /θ/ differently, when they both have /f/ and /s/ natively?

The present thesis only focused on the voiceless fricatives of the target languages, as neither Mandarin nor Cantonese has voicing as an active phonological feature in their inventories. The glottal fricative /h/ was also excluded from any discussions on fricatives in this thesis, as it was considered fundamentally different from other fricatives, produced without any articulatory constriction within the oral cavity but in the pharynx.

Table 1-1 A comparison of English, Mandarin and Cantonese inventories of fricative phonemes, sorted by their places of articulation (Bauer & Benedict, 1997; Cheng, 1973; Duanmu, 2007; Giegerich, 1992; Xu, 1989). Phoneme symbols with brackets means that there were disagreements on whether they were phonemes or allophones of other phonemes.

	labio-dental	**dental**	**alveolar**	**palatal-alveolar**	**retroflex**	**alveolo-palatal**	**velar**	**glottal**
English	f v	θ ð	s z	ʃ ʒ				h
Mandarin	f		s		ʂ (ʐ)	(ɕ)	x	
Cantonese	f		s					h

1.1 Second-language speech sound perception

Evidence has shown that speech sounds are perceived categorically instead of continuously, and each sound category has a prototype serving as a perceptual magnet, pulling perceptually similar sounds into the category (Kuhl, 1991; Liberman et al., 1957). These sound categories are specifically developed from L1 input, and enable more efficient L1 processing (Cutler, 2000; Flege, 2002). Forming L1-specific phoneme categories supports processing of native speech, as only acoustic information that is helpful for differentiating categories is maximized for perception while differences within a category are minimized (e.g. Kuhl et al., 2006). This results in economic cognitive effort distribution and increased L1 processing accuracy. While developing categorical perception of L1 sounds, a set of speech processing strategies is also developed. Based on this approach, what acoustic information within a speech sound is deemed important is highly language-specific. The same acoustic signal may be weighted differently by two listeners, depending on their listening strategies modified by their language experience (e.g. Iverson, Hazan, & Bannister, 2005; Iverson et al., 2003).

Language-specific phoneme categories may pose challenges to processing non-native speech sounds. One's listening strategy is highly tuned for native sound contrasts through years of learning and adapting, and thus less sensitive to some L2 sound contrasts (Kuhl et al., 2006). This is because listeners are trained to attend to certain acoustic cues that are crucial, and ignore cues that are less important for differentiating

native phoneme categories; however, the cues they are trained to ignore might be crucial for identifying some L2 phoneme categories (e.g. Iverson et al., 2003).

Research has attempted to model the interaction between L2 sounds and L1 phoneme categories. Best (1995, p. 193) claims that L2 sounds tend to be perceived according to their similarities and dissimilarities to the native phoneme categories that are "in closest proximity to" the L2 sounds in one's phonological space. The perceptual assimilation model (PAM) was proposed based on this theory. PAM points out that L2 segments that are considered as speech sounds by L1 listeners will be either assimilated into a L1 phoneme category based on its place and manner of articulation, or uncategorised and fall in between categories. L2 listening difficulty arises when a pair of L2 sounds is assimilated into the same L1 category, referred to by PAM as "single category assimilation", since the L2 contrast is neutralised by listener, and the discrimination of the contrast is expected to be poor. Another model proposed around the same time is Flege's speech learning model (SLM) (Flege, 1995). Similar to PAM, SLM also proposes that perceptual difficulties may happen when a learner fails to establish a separate category for an L2 sound, but instead processes it as an L1 category. Category formation of an L2 sound may be impeded by "the mechanism of equivalence classification" (Flege, 1995, p. 239); as a result, one category is used to process both sounds, and distinctions are overlooked. Different from PAM, SLM points out that comparisons between L2 sounds and L1 categories during their interaction take place at a phonetic level instead of a gestural level, assuming more detailed acoustic differences rather than gestural similarities are at play. PAM-L2 has bridged this discrepancy, as it argues that L1 listeners with some L2 learning experience may be influenced by both gestural and phonetic details, and other phonological factors (Best & Tyler, 2007). The same L2 category may be assimilated to different L1 categories at different times, depending on acoustic factors like allophonic features, syllable position and stress. In other words, allophones of the same phoneme category of L2 may be treated differently by the L1 system due to their different phonological systems. One may also deduce that sound categories that are transcribed phonemically with the same symbol by different language can have distinct prototypes, leading to distinct results of categorising an L2 sound.

In summary, L1 experience has an undeniable influence on L2 sound processing. By combining knowledge about language-specific cue weighting strategy and L1-L2 phonological interaction, it becomes clear that, to understand an L2 sound assimilation pattern, it is crucial to understand the phonetic features of relevant L1 and L1 sound categories, and the detailed cue weighting system of listeners.

One common assumption of those L1-L2 interaction models is that L1 listening strategy and L2 listening strategy are the same for a listener. In other words, instead of developing a new set of strategies for processing L2 sounds, listeners tend to use one set of processing strategies to perceive all sounds. The assumption was tested in a cross-linguistic comparison study investigating perception of coarticulation (Wagner et al., 2006) which discovered that the L1 listening strategy was also applied when processing L2 sounds. In the study, Spanish-speaking subjects adopted the same identification strategy when listening to both Spanish and Dutch fricatives, specifically paying attention to cues that native Dutch listeners would not. They concluded that fricative processing strategies appeared to vary according to the native fricative inventory, and it had an impact on the strategy for identifying non-native fricatives. It was clear that the listeners use their L1 listening strategies when processing non-native sounds.

Evidence has shown that this set of strategies may be constantly changing due to the listening experience, specifically the amount of exposure to L2. This is a consensus shared by SLM and PAM-L2 (Best & Tyler, 2007; Flege, 1995). As Flege (1995) mentioned, when L2 sounds are introduced to an L1 phonological system, the effect is bi-directional, and the end result would be L1 and L2 sounds co-existing in one perceptual space. Native phoneme categories and listening strategies are shaped and reinforced by abundant L1 input; meanwhile, extensive L2 exposure may also shake, even reform, a developed system to some extent. In this scenario, L2 learners may develop a fusion of L1 and L2 strategies that works for phonological systems of both languages. Evidence for this was provided by a study by Wang, Behne, and Jiang (2008). The study discovered that Mandarin-speaking subjects with around 10 years of Canadian residency demonstrated task performance approximating the native English speakers, while the subjects with no more than 4 years of residency demonstrated poorer identification of English-specific sounds. This study revealed that

L2 learners could achieve native-like identification accuracy with years of training. Another study (Chang, 2010) suggested that changes may take place even earlier. It investigated the L1 phonetic development during language contact, i.e. learning an L2, and it discovered an L1 phonetic drift that started from the onset of L2 learning. The study argued that there was significant modification in phonetic representations in the L1 phonological system from as early as 2 weeks. However, this study focused mainly on changes in production, and whether it is necessarily the case that perceptual changes also took place so early in learning was not discussed. Nonetheless, it is clear, when discussing the interaction between L1 and L2 sound systems, that one should not neglect the effect of an L2 on a listener's processing strategy.

1.2 Fricative acoustics and perception

Whether a fricative is from an L1 or L2, sources of acoustic cues for perceiving it are the same, as they primarily derive from the inherent articulatory mechanisms used to produce them. These mechanisms always involve narrow turbulence-producing constrictions formed using the articulators in the vocal tract (Shadle, 1990; Stevens, 2000). The most significant production parameters of fricatives are the presence of an obstacle (e.g. the upper teeth when producing /s/), the length beyond the constriction, and the flowrate of the airstream (Shadle, 1985). These parameters lead to various measurable acoustic features in the frication noise. The features of frication can be characterised by energy distribution, noise amplitude, and noise duration (Jongman et al., 2000).

When fricatives are produced in a sequence of phonemes (i.e. syllables or sentences), coarticulation effects occur, giving another source of acoustic cues for fricative identification. Coarticulation effects takes place in between two neighbouring sounds, which makes coarticulation contain some articulatory features from both sounds. In coarticulation, the acoustic features of a segment are extended beyond its boundary, and available to perception "longer than would be the case if all cues were confined inside its boundaries" (Kühnert & Nolan, 2009, p. 62). In the case of a syllable with fricative as its initial consonant, coarticulation effects mean that the early portion of the vocalic section carry some information from the fricative as well (Wilde, 1995).

As well as frication noise and coarticulatory effects, the identification of a fricative may also be affected by phonological cues (e.g., phonotactic information and phonological rules, Weber & Cutler, 2006), and visual cues (Strand & Johnson, 1996; Wang et al., 2008). The present focuses are frication and coarticulation cues, and the role they play in fricative perception.

There is no question about the primary status of the frication noise as a source of cues for fricative identification. Acoustically, spectra of fricatives contain a sufficient amount of information to distinguish places of articulation. Indeed, spectral properties of fricatives alone could distinguish sibilants from non-sibilants, and all fricatives could be distinguished with any two properties. (Jongman et al., 2000; Stevens, 2000). Perceptually, frication also appears to be sufficient for identifying fricatives accurately in some studies. The study by Borzone de Manrique and Massone (1981) investigated the perception of Argentine Spanish fricatives /f s ʃ x/ by asking native subjects to repeat stimuli they heard. These were all lengthened fricatives with frication alone. The results showed that with untreated stimuli, the subjects reached an accuracy level of above 95% for all the fricatives. Zeng and Turner (1990) also provided evidence showing near 100% accuracy at identifying English fricatives with frication presented alone at and above 50 dB.

The role of coarticulatory cues in fricative identification is less clear. Coarticulatory effects are common to all phoneme types, and there was evidence showing that formant transition patterns contain information like place of articulation (Wilde, 1995). a controversy about what role they play surfaced when comparing the studies of coarticulatory cues for fricative identification. Some studies (Behrens & Blumstein, 1988b; Harris, 1958; Nittrouer & Studdert-Kennedy, 1987; Nittrouer, 1986; Strevens, 1960) pointed out that the use of formant cues is fricative-specific, as the fricatives that are acoustically distinct could distinguish themselves from other fricatives with frication alone. For example, English /s/, which shows a distinct spectral peak and substantially larger noise energy than English /f/ and /θ/, was considered perceptually distinct with just its frication. Meanwhile, fricatives that are less spectrally distinguishable would require extra information, like formant transitional cues, for accurate identification. However, this selective dependency on coarticulatory cues was shown to develop through age (Nittrouer & Studdert-Kennedy, 1987; Nittrouer &

Miller, 1997); in other words, this listening strategy was cultivated by years of language experience. Some other studies thus argued that the use of formant cues for fricative perception is language-specific instead of fricative-specific (Wagner, 2013; Wagner et al., 2006; Wagner & Ernestus, 2004), and was strongly associated with the native fricative inventory.

The study by Wagner et al. (2006) offered a lot of evidence supporting the latter view. They argued that there may not be a universally distinct fricative, as distinctness should be language-specific. They conducted a number of experiments with listeners from different language backgrounds to evaluate the importance of formant transitional information and L1 knowledge. The studied languages—Spanish, English, Dutch, and German—had various sizes of fricative inventories. Among these languages, German and Dutch only have spectrally distinct fricatives, while the other languages have pairs of fricatives that are spectrally similar. This study used a phoneme monitoring task with stimuli that had either matching or mismatching formant transitions from fricative to vowel. The results showed language-specific patterns in the use of transitional cues for fricative identification; German and Dutch speakers were not affected by mismatching transitions, while the performance of the speakers of the other language groups was affected, demonstrated by lower accuracy and longer reaction time. Moreover, different listeners found different fricatives harder to identify, depending on their native language background. For example, it appeared that Spanish speakers paid attention to formant transitions when identifying /f/, and Polish speakers paid attention to them when identifying /s/. Apart from revealing language-specific listening strategies for fricative identification, it also showed that the perceptual robustness of fricatives is likely also be language-specific. This means that whether speakers use transitional cues may not depend on the inherent distinctiveness of the fricatives, but on whether there is a similar category in their native fricative inventory. Indeed, in another study Wagner (2013) followed up these findings, and confirmed that listeners whose native language contained acoustically similar fricatives would make use of more than just frication noise as an information source in order to increase their identification accuracy. This finding remained true for fricatives in both syllable-initial and syllable-final positions.

Interestingly, coarticulatory cues seem to interact with frication cues: when both are present, the weight assigned to each differs from when only one type of cue exists. This was observed by Borzone de Manrique and Massone (1981). The study presented native Argentine Spanish listeners naturally produced fricatives (in isolation and within syllables), synthetic fricatives in isolation, and concatenated naturally produced fricatives and vowels (transitionless), and the listeners identified and labelled the fricative segments. The results indicated that, even though the frication portion alone was sufficient for identifying almost all Argentine Spanish fricatives, natural or synthetic, transitionless syllables impeded accurate identification. In this case, it appears that transition cues shared the perceptual weight on frication in a context where both cues are, or should be present. Moreover, more perceptual weight was put on cues within coarticulation when frication noise is presented in a degraded form. When the speech-to-noise ratio decreased, or the intensity level of the frication was reduced, the presence of coarticulatory cues enhanced the accuracy of fricative identification (Alwan et al., 2011; Zeng & Turner, 1990). These findings seem to suggest that the use of coarticulatory cues may not be language-specific or fricative-specific, but instead is essential for fricative perception in degraded listening conditions—such as a low speech-to-noise ratio or when speech is quiet—which could lead to low intelligibility of frication. When the listening conditions are less than ideal, listeners, in spite of their language background, likely rely more on transitional cues.

One may argue that identifying an L2 fricative could be considered as one of these less-than-ideal listening conditions. Admittedly, these findings presented in the previous paragraph were based on data collected from Spanish and English speaking subjects. According to Wagner et al. (2006) and Wagner (2013), because Spanish and English have rich fricative inventories with perceptually similar fricatives, turning to coarticulation for cues is part of their speakers' listening strategy. Whether listeners whose native fricatives are all distinct from each other (e.g., Dutch) will also turn to coarticulatory cues for information under a degraded or a non-native listening condition is unclear. Research has speculated that Dutch listeners do use coarticulatory cues when perceiving English /θ/, as they assimilate it to their native /s/ category whose frication is not acoustically close to /θ/ (Johnson & Babel, 2010). However, this needs further research.

To sum up, while the majority of perceptual weight is put on frication cues, the role of coarticulatory cues such as formant transitions in the perception of fricatives is still unclear. Nevertheless, it is clear that studying fricative perception should not leave out coarticulation. The perceptual weight on coarticulation may be determined by various factors, including listening conditions and language background. Coarticulatory cues may play different roles in different languages, depending on the availability of an acoustically similar category in a native fricative inventory. The perception of coarticulation appeared to be a crucial factor in determining the L1 listening strategy for fricatives, and consequently, it may result in differential perception of a L2 fricative category.

1.3 Cantonese and Mandarin fricatives

It is clear that their differential assimilation of English /θ/ between Cantonese and Mandarin native speakers is not driven solely by what is available in their fricative inventories, as both Mandarin and Cantonese have /f/ and /s/. Therefore, it is important to have an overall understanding of both languages, and observe what other factors may be playing a role in their perception of /θ/.

Mandarin (sometimes referred to as standard spoken Chinese, or Putonghua in studies) is a standardized Chinese dialect, with its standard pronunciation developed based on the Beijing dialect. Mandarin is used as the lingua franca of mainland China and other Chinese communities around the world (Duanmu, 2007). Cantonese, which used to function as the standard spoken language of Canton province, is one of the major varieties of the Chinese language. It is mostly spoken in southeast China, and still is used as the lingua franca in Hong Kong (Bauer & Benedict, 1997; Chan & Li, 2000; Ramsey, 1989). Cantonese and Mandarin have differences in many aspects, including in their phonology and phonetics. As shown in Table 1-1, Mandarin has a larger fricative inventory than Cantonese which only possesses two fricative phonemes. This contrast may lead to differences in both acoustics and perception of the fricative categories, even the ones that are common to the two languages.

As discussed in section 1.1, even though Cantonese and Mandarin both contain /f/ and /s/, they may be spectrally different in each language. These acoustic differences could

lead to different weight assignment of cues within frication, and different dependence on coarticulatory cues by listeners of various language backgrounds. This section is going to outline research investigating Cantonese and Mandarin fricatives, both in terms of their perception and their acoustics.

1.3.1 Cantonese fricatives

There is a limited number of studies on the acoustic features or the perception of Cantonese fricatives, likely due to the small number of native fricatives in Cantonese: only one sibilant, /s/, and one non-sibilant, /f/. Most of the studies reviewed in this section discuss the Cantonese phonological system more generally but also include some information on fricatives.

There are some controversies regarding the production of /s/ in Cantonese. A chronological merger of /s/ and /ʃ/ in Cantonese has been reported by some researchers (e.g., Cheung, Gan, & Zhan, 1987; Hashimoto, 1972; Pulleyblank, 1997), with [ʃ] and [s] existing as allophones of a single category, /s/. Some have argued that they are in free variation (Pulleyblank, 1997), while others have suggested that they are in complementary distribution depending on the vowel contexts. Additionally, [ɕ] is thought to be in free variation with [ʃ] for some speakers (Bauer & Benedict, 1997; Hashimoto, 1972). Despite the debate over how the variants are distributed, there is little dispute about the neutralisation of the contrast between /s/ and /ʃ/, and that this has led to a wider range of variants within the /s/ category. Articulatory features of Cantonese /s/ were provided by some other studies. The study by Zee and Xu (1999) argued that in Cantonese /s/ the tongue tip is extended further to reach the teeth in order to maximize the contrast with /ʃ/ which used to be a separate phoneme category. This argument was supported by Lee's study (1999), in which palatographic images were obtained from subjects who were asked to read Cantonese monosyllabic words with /s/ as syllable initial, followed by vowels /i y ɛ a ɔ/. The palatographs demonstrated that Cantonese /s/ was lamino-alveolar, involving the blade instead of the apex of the tongue in articulation. The constriction formed ranged from the upper portion of the front incisors to the alveolar region.

A later study by Lee and Zee (2010) again obtained palatograms of native Cantonese speakers producing /s/-initial monosyllables, with vowels /i y a ɔ/, to investigate its

articulatory characteristics. This study confirmed the findings of the earlier study by Lee (1999), showing that Cantonese /s/ was produced with lamino-alveolar articulation. However, different from the previous studies, Lee and Zee (2010) discovered that, though the production of /s/ varied systematically according to the vocalic contexts, the constriction region did not change much. Instead, the different variants arose as a result of changes in the size of the constriction area. This finding was supported by an acoustic study by Yu (2016), which analysed the spectral, amplitude and duration features of Cantonese /s/ in various vocalic contexts (/i y ɛ a ɔ/). The measures suggested that the acoustic properties of Cantonese /s/ were more consistent with the properties of /s/ rather than the /ʃ/ or /ɕ/ of other languages. In addition, this study also showed that there was significant "inter-individual variability", and that this was caused by features of the vowel. For example, some subjects' production of /s/ was more affected by vowel height, while others were more affected by lip-rounding.

There are no detailed acoustic or perceptual studies of Cantonese /f/. However, two features were mentioned in some studies: firstly, lip-rounding of /f/ when it is followed by rounded vowels (Bauer & Benedict, 1997); secondly, that /f/ occurs less frequently than Cantonese /s/ in speech (Leung et al., 2004).

1.3.2 Mandarin fricatives

Mandarin has a larger fricative inventory than Cantonese (as shown in Table 1-1). Other than /f/ and /s/ which are shared by the two languages, Mandarin also has /ʂ/, /ɕ/ and /x/. As a result, the potentially more complicated interactions within the perceptual space have led to more studies investigating the physical features and the perception of Mandarin fricatives. The study by Lee, Zhang and Li (2014), replicated the analyses used by Jongman, Wayland, and Wong (2000) in their study of English fricatives), and is so far the most thorough acoustic study of fricatives in Mandarin. They recorded 6 Beijing Mandarin speakers saying Mandarin CV syllables with initial fricative followed by the vowel /a/, and reported 11 acoustic measures for all Mandarin fricatives (except the voiceless velar fricative /x/) that analysed spectral, amplitude, and duration information. Their results showed that no single measure could classify all Mandarin fricatives, and that at least two measures were needed to distinguish all

fricative contrasts. In addition, no single fricative was distinct in every measure. The fricative pairs /f/-/ɕ/ and /f/-/ʂ/ could be distinguished by the highest number of measures (7), and /f/-/s/ could also be classified by 5 measures. The study concluded that /f/ as the only non-sibilant in the analysis was the most distinct fricative in Mandarin. On the other hand, the measures were less effective for distinguishing sibilants. The most confusing fricative pair was /s/ and /ɕ/, which could only be distinguished by 2 measures. The study also compared fricatives of English and Mandarin, and concluded that the most effective measures for classifying fricative place also differed between the two languages: spectral skewness and F2 onset were the most effective features for Mandarin, while spectral skewness, variance, and relative amplitude were the most effective features for English. The indication of this finding could be that English and Mandarin native speakers may rely on different acoustic cues to identify fricatives, or may rely on the same cue to different extents. The most effective measures for Mandarin fricatives appeared to be spectral skewness and F2 frequency at vowel onset, as each could distinguish 5 fricative pairs in their study.

Instead of limiting the vowel contexts of the fricative syllables, Svantesson and Shi (1986) conducted an acoustic analysis on all the Mandarin fricatives by measuring the production of all the possible combinations of target fricatives and vowels in Mandarin. Four male speakers of Mandarin (which they referred to as "standard Chinese") were recruited from northern parts of China, including Beijing and Liaoning. They observed the shape of the waveforms of the fricatives, and in order to characterize the spectral shapes of the target sounds, measured the centre of gravity, the dispersion of spectra, and the mean intensity level. After plotting the CoG values against the dispersion measurements, they demonstrated that Mandarin /f/ was separated from all other fricatives by having a flatter spectrum and higher dispersion, while sibilants were with low dispersion and high CoG. If a diagram was plotted representing similarity and dissimilarity among Mandarin fricatives defined by the CoG and dispersion plot, it would look like Figure 1-1. The fricatives were clearly grouped into sibilants and non-sibilants. This study did not make any further inferences from the crowding of sibilants demonstrated by the figure. However, it could mean that in order to maximize contrast, Mandarin speakers use more extreme articulatory positions in an attempt to separate the sibilants.

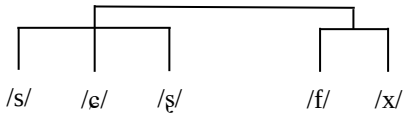


Figure 1-1 Diagram representing the similarity and dissimilarity among Mandarin fricatives based on their spectral properties of the production data from Svantesson and Shi's study (1986).

One study by Zhang, Lü, and Qi (1982) analysed the perceptual similarities of Mandarin fricatives for native speakers, and the diagram they proposed based on perception was different from that shown in Figure 1-1. Similar to the study by Svantesson and Shi (1986), they only tested participants native to Beijing using stimuli produced by speakers from Beijing. The 16 participants completed an identification task, in which they wrote down the syllables presented in different transmission conditions including various filtering processes and speech-to-noise ratios. The study used cluster analysis to explore the perceptual confusions of consonants including fricatives. They proposed a perceptual confusion diagram based on Mandarin phoneme categories (vowels, stops, and fricatives etc.) and perceptual similarities within each category, and the diagram for Mandarin fricatives is shown in Figure 1-2. When comparing the two figures, it appears that /ɕ/ is grouped differently by the two studies. Acoustically, /ɕ/ was characterized as a sibilant which was similar to /s/ and /ʂ/. However, perceptually, /ɕ/ was distinct from other sibilants and non-sibilants. This was likely due to the near complementary distribution of /ɕ/ and /s/ or /ʂ/ (Cheng, 1973; Duanmu, 2007; Svantesson & Shi, 1986). To be specific, /ɕ/ is the only Mandarin fricative that can be followed by close front vowels /i/ and /y/. As /ɕ/ almost never occurs in the same vowel contexts as /s/ or /ʂ/, the perceptual dependence on fricative cues may be reduced, making /ɕ/ perceptually distinct from other Mandarin fricatives despite its acoustic similarities with other sibilants.

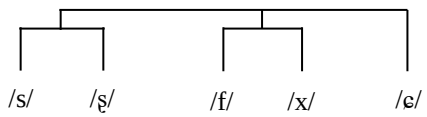


Figure 1-2 Diagram representing perceptual similarities of the Mandarin fricatives extracted from the study by Zhang et al. (1982).

Other than the different grouping of /ɕ/, the two studies discussed above are in agreement on the relationship between other sibilants and non-sibilants of Mandarin. To be specific, they considered the fricative pairs— /s/ and /ʂ/, and /f/ and /x/— to be acoustically and perceptually similar. Another study on the acoustics of Mandarin

fricatives updated that view with another production analysis using different parameters. Li, Edwards, and Beckman (2007) analysed the spectral properties of sibilants in Mandarin, and compared them to English and Japanese. They measured two spectral parameters in their study: amplitude ratio, which assesses the degree of palatalization, and centroid frequency above the F2 region, which aimed to reveal the place of articulation. The results revealed that Mandarin /ʂ/ differed from other Mandarin sibilants in both spectral parameters, meaning that either of the two parameters alone could distinguish /ʂ/ from the other sounds. However, /s/ and /ɕ/ required both parameters to be distinguished. When compared to English, Mandarin /s/ appeared to differ: the English /s/ had lower centroid frequency in the high frequency band. They concluded that the position of the narrowest constriction formed to produce Mandarin /s/ is more to the back of the oral cavity compared to English /s/. This result also led to the conclusion that phoneme categories of different languages which were transcribed with the same symbol could have very distinct acoustic features and category prototypes.

A similar conclusion was drawn by Proctor, Hsuan, Lu, Zhu, Goldstein, and Narayanan (2012), though based on an opposite result. The study measured the place and manner of articulation of Mandarin sibilants using real-time Magnetic Resonance Imaging (rtMRI). They recruited Mandarin speakers from a variety of places in southern China. The results showed that /s/ was produced either as an alveolar fricative or a dental fricative in Mandarin, and the authors concluded that Mandarin /s/ has a highly apical and anterior coronal fricative, with some speakers moving their tongues to the dental place of articulation. This could be one of the factors that distinguishes Mandarin /s/ from /s/s in other languages. In addition, they discovered that the retroflex feature of /ʂ/ disappeared in some speakers and that they tended to produce /ʃ/ instead, with the point of maximal constriction ranging from alveolar to palatal. This shift was not reported in studies of northern Mandarin fricatives (Svantesson & Shi, 1986; Zhang et al., 1982), so whether it is also taking place in northern China is unclear. Compared to northern China, southern China has a more complex dialectal environment, and Mandarin is more likely not the (only) native language of people from the south. This indicated that studies focussing on features of Mandarin should control for participants' language experience by limiting recruitment to one area with participants from similar language environments.

To sum up, Mandarin fricatives can be grouped into sibilants and non-sibilants, both perceptually and acoustically, with the exception of /ɕ/ which is acoustically a sibilant but is perceptually distinct from other sibilants and non-sibilants. Amongst sibilants, /ʂ/ appears to be relatively more acoustically salient, and the only phoneme /s/ that is present in both English and Mandarin is reported to be acoustically different in these two languages.

1.3.3 Comparisons between Cantonese and Mandarin fricatives

Cantonese and Mandarin fricatives also share some features. First of all, they can only occupy the initial position of syllables, and never appear in consonant clusters (Bauer & Benedict, 1997; Duanmu, 2007). Another common feature is that their assimilation patterns of English /θ/ appeared to be consistent despite where /θ/ was in an English syllable. Peng and Setter (2000) studied the assimilation of /θ/ at various syllable position, and showed that /θ/ was frequently replaced by /f/ by Cantonese native speakers regardless of whether it was in the syllable or syllable coda. Jiang (1995) claimed that Mandarin native speakers also consistently replace /θ/ with their /s/ category despite the different positions /θ/ may occupy. One may conclude that the differential assimilation between Cantonese and Mandarin was not affected by the syllable structure of the input, but was more likely affected by the different acoustics and/or the perception of the fricatives.

In terms of articulation, Stokes and Fang (1998) conducted an electropalatographic (EPG) analysis of Mandarin sibilants, and compared it to existing descriptions of Cantonese and English from other studies. They discovered that Mandarin /s/ was characterized by a narrow anterior groove, little contact in the mid-palate, and a firm lateral seal, which appeared to be very similar to the EPG image of English /s/ provided by Hardcastle, Gibbon, and Jones (1991). On the other hand, the EPG image of Cantonese /s/ was different from the /s/ of Mandarin and English; Cantonese /s/ resembled Mandarin /ɕ/'s contact pattern that had slightly bigger mid-palatal contact. This contrasts with the results of an acoustic analysis of Cantonese /s/ (Yu, 2016), which states that Cantonese /s/ is more similar to other languages' /s/ than to /ɕ/. To resolve this contradiction, more acoustic and perceptual studies on Cantonese fricatives are required.

It is also crucial to explore what acoustic differences are led to by differences in articulation. Some acoustic information was provided by Li (2018), as they analysed the production features of Mandarin sibilants produced by late Cantonese-Mandarin bilinguals from Cantonese-speaking communities. Cantonese was their dominant language, and they did not start to learn Mandarin until their 20s. The results revealed that the /s/ produced by these late bilinguals was significantly different from the /s/ produced by native Mandarin speakers in terms of the spectral peak location and the F2 frequency at vowel onset. The late bilinguals' production of Mandarin /s/ had a lower spectral peak and higher F2 onset, similar to their Mandarin /ʂ/, indicating that they did not have a clear categorical distinction when producing the two categories. They suggested that the reason for such a phenomenon may be that Cantonese /s/ was acoustically midway between Mandarin /s/ and /ʂ/.

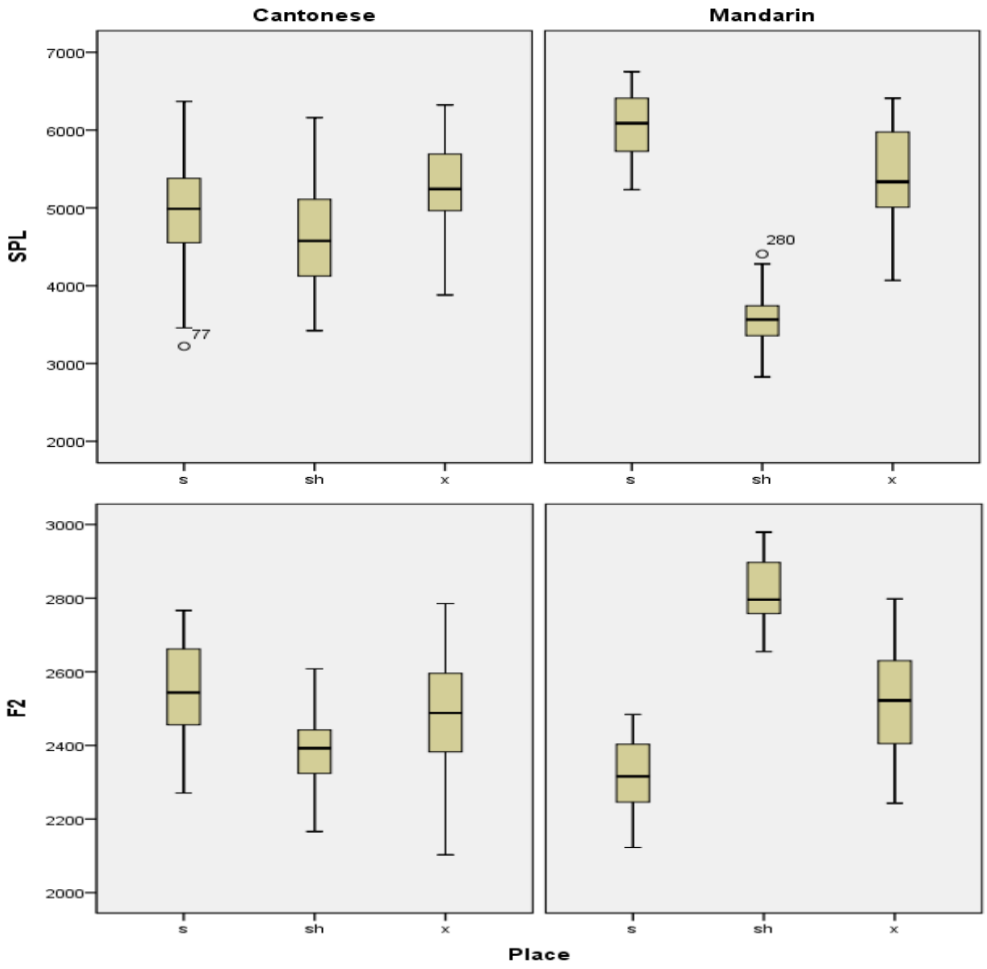


Figure 1-3 Spectral peak location (SPL) and F2 onset (F2) values (in Hz) of /s/, /ʂ/, and /ɕ/ (represented by Pinyin s, sh, and x respectively) produced by Cantonese and Mandarin native speakers.
Note. This figure was reproduced from "The Production of Mandarin Voiceless Sibilant Fricatives by Late Cantonese- Mandarin Bilinguals: An Acoustic Study" by Y. Li, 2018, English Literature and Language Review, 4(5), p. 83. Copyright 2018 by Y. Li.

A study of both the perception and production of fricatives by Zheng and Iverson (2016) included an assimilation test and an acoustic analysis to investigate how the production of fricatives affected Mandarin and Cantonese speakers' assimilation of English /θ/. They tested subjects with very limited or no experience of English living in northern China and Hong Kong. The subjects read out real words in their native language, and accomplished an assimilation test. The acoustic analysis demonstrated that Cantonese /s/ and Mandarin /s/ had very similar spectral peak locations, while the CoG value of Cantonese /s/ was lower than for Mandarin. In addition, the study revealed a difference in spectral shape between the /f/ categories; Mandarin /f/ appeared to be more peaked than Cantonese /f/. The study also compared /s/ and /f/ categories of Mandarin and Cantonese to English /θ/, and found that while Cantonese /f/ and English /θ/ were spectrally similar to each other, Mandarin /s/ and /θ/ were distinct. The result of the assimilation test, shown in Figure 1-4, revealed that the Cantonese subjects perceived nearly all the exemplars of English /θ/ as their /f/ category, while the Mandarin subjects perceived slightly more than 70% of the /θ/ tokens as Mandarin /s/, and around 30% as Mandarin /f/. Although the overall result fits the conclusion of previous studies (Jiang, 1995; Hung, 2000; Chan & Li, 2000; Peng & Setter, 2000), it showed that though the Mandarin speakers preferred their native /s/ category to the /f/ category when it comes to assimilating /θ/, their preference was not as strong as the preference of the Cantonese speakers. More importantly, the acoustic features revealed in this study could not predict the assimilation patterns of /θ/ by the Mandarin listeners.

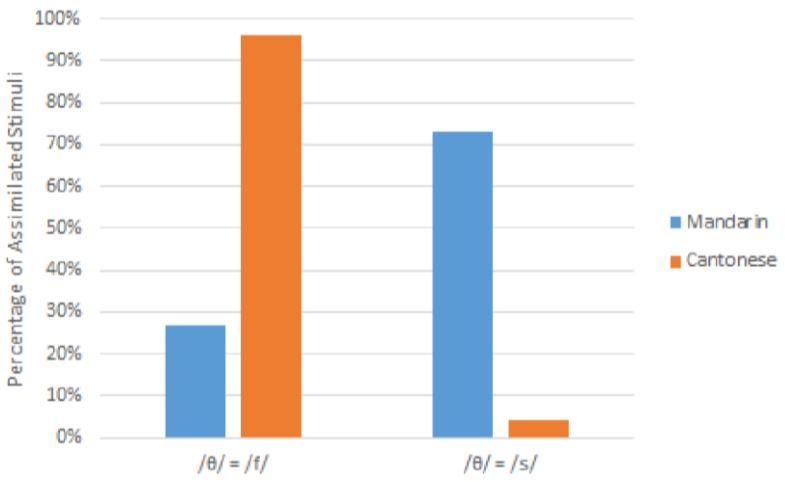


Figure 1-4 The result of the assimilation test of English /θ/ by Mandarin and Cantonese native speakers (Zheng & Iverson, 2016).

In summary, the results of the studies reviewed above have led to a conclusion that even though /f/ and /s/ exist in both Cantonese and Mandarin, they differ in terms of their acoustic properties. However, there is disagreement as to how they differ and it was still unclear how such differences lead to differences in perception of /θ/ by Mandarin and Cantonese listeners.

1.4 Thesis Overview

The current project aims to investigate the differential assimilation of English /θ/ by Cantonese and Mandarin native speakers, using multiple experiments to investigate the acoustic properties of relevant voiceless fricatives, and the interactions between L1 and L2 sound processing strategies. Three studies are reported in separate chapters.

Chapter 2 reports details of Study 1, which explored the acoustic properties of target fricatives of Cantonese, Mandarin, and English, and perceptual weighting of the acoustic cues, using speech stream recordings from native speakers. The chapter reports results from a traditional acoustic analysis and a novel machine-learning approach; both methods compare fricatives within and across languages. The within-language fricative comparisons are used to explore the effectiveness of each cue in terms of their efficiency in distinguishing native fricative pairs, based on which the native perceptual weight of the cues was estimated. The cross-language comparison aimed to reveal the motivations for the different assimilation patterns of English /θ/ by Cantonese and Mandarin speakers. The chapter discusses findings about the fricative categories that are shared by three languages, and argues that the differences in cues may influence perception, revealing an L1 effect. The results also demonstrate how various acoustic cues show different levels of efficiency in distinguishing fricatives, indicating that native speakers of these languages may rely on these cues differently.

Chapter 3 presents Study 2, which focused on the perception of formant transitions in fricative syllables. The study examined how transition cues caused by coarticulation affect the perception and categorisation of /f/ and /s/ by native Cantonese, Mandarin, and English speakers. Previous work has suggested that speakers of languages with small fricative inventories, like Cantonese, are much less dependent on formant transitions, likely because they can distinguish among their native-language fricatives

using frication cues alone. Based on this, Study 2 formed the hypothesis that different levels of dependency on transition cues may be the reason for the different assimilation of /θ/. This study investigated this question through an integration of a behavioural experiment and electroencephalogram (EEG) measurements into an active oddball paradigm. The behavioural component was a phoneme monitoring task, and from the EEG recordings P300/P3b potential (referred to as P300 in this thesis) was measured. The stimuli were fricative syllables spliced with vowels such that the formant transitions were either congruent or incongruent with the fricatives. The paradigm is used to investigate whether listeners attend to formant transitions when identifying fricatives or not. The results are discussed with reference to listeners' differing language backgrounds and fricative inventories.

Chapter 4 details Study 3, which investigated cross-language differences in categorisation boundaries of target fricative pairs, aiming to discover how different fricative inventories affect fricative perceptual space. The study used a behavioural phoneme identification task, in which subjects labelled syllable-initial fricatives. The stimuli were created by interpolating the fricative pairs and generating a frication continuum and a vowel continuum, forming a 2-dimensional stimuli grid. A perceptual boundary was located on each stimuli grid based on the results of the identification task. This chapter discusses whether frication or vocalic information is the primary source of cues for fricative identification for the native English, Cantonese, and Mandarin speakers. Results are examined to investigate if any cross-language differences in fricative boundaries uncovered in this study could address the main research question.

Chapter 5 presents a general discussion that summarises the key findings of the studies and interpretations of the findings. Implications of these findings are also presented in this chapter in terms of how they conform with or challenge the existing theories on L2 fricative perception.

Chapter 2 Fricative Properties: Acoustic Cues Analysis

2.1 Introduction

To understand the processing of fricatives and weighting of acoustic cues, it was essential to have a whole picture of what cues are available within the target fricatives of each language. Much of our knowledge of fricative acoustics has come from analyses of recordings of CV structured syllables, or mono-syllabic words, even though fricatives, or any speech sound processing in real life, would have to deal with them embedded in speech streams that may involve complex interactions among sounds. This view inspired the study that would be introduced in this chapter which shifted the focus of acoustic analysis to sounds from speech stream, combining some traditional and novel acoustic measurements proposed by previous studies. In addition, with the intention to answer the research question stated in the previous chapter, this study was the first attempt to analyse and compare fricative acoustics of Cantonese, English, and Mandarin in the same study.

In terms of what acoustic features should be informative for fricative perception, the following sections referenced the studies by Jongman et al. (2000) and Lee et al. (2014) as guidelines. A comparison between the two studies could provide an idea on how each type of acoustic measures would classify fricatives differently cross-linguistically, and based on this comparison, inferences on how those measures may affect fricative perception could be made. Three types of measurements were reported in the two studies—spectral (including frication and transition), amplitude, and duration—each had different impacts on perception. There was not a study analysing the acoustic properties of Cantonese fricatives using similar measurements likely because Cantonese has a small fricative inventory. For the purpose of the present study, only findings about voiceless fricatives were discussed.

2.1.1 Fricative spectral property measurements

Studying fricative spectra would provide insight for the narrow constriction formed in the oral cavity, as they are determined by the shape and length of the cavity anterior to the constriction. Fricative spectral properties included measurements of spectral peak

location, four spectral moments, and transitional information including locus equation and F2 at vowel onset.

Previous studies on English fricatives have discovered that the peak location of a spectrum could distinguish not only sibilants from non-sibilants, but also sibilant fricatives from each other (e.g. Heinz & Stevens, 1961; Stevens, 2000). Jongman et al.'s extensive acoustic analysis (2000) reported that the four places of articulation of English fricatives were significantly different from each other. On the other hand, a study on Mandarin fricatives revealed that their spectral peak values were sufficient to distinguish /ʂ/ from the other fricatives, but not sufficient to distinguish /f/, /s/, and /ɕ/ from each other (Lee et al., 2014). We may infer from this contrast that spectral peak location as an acoustic property may play different roles in fricative perception of different languages, depending on how efficiently it could distinguish fricatives within a native inventory. It is possible that Mandarin listeners rely less on the spectral peak location than English listeners, and this difference may be one of the factors which lead to their categorisation of /θ/. Therefore, it is highly necessary to investigate this spectral feature.

Admittedly, fricative spectra are not characterized by a single peak (Hughes & Halle, 1956; Stevens, 2000). However, analysing only the most prominent peak of would allow a comparison among studies, as most acoustic analyses of fricatives were only discussing the highest point of a spectrum, as the main peak was sufficient to characterize many fricatives, especially sibilants (Behrens & Blumstein, 1988b; Gordon et al., 2002; Jongman et al., 2000). The same decision was made by Lee et al. (2014) for their acoustic analysis of Mandarin fricatives.

Spectral moments analysis aimed to capture energy concentration and distribution of fricatives. The four spectral moments were mean, variance, skewness, and kurtosis: the first two measurements could reflect spectral energy concentration and range, and the latter two could reveal details of energy distribution, including distribution asymmetry and peakedness (Forrest et al., 1988; Jongman et al., 2000). Both Jongman et al. (2000) and Lee et al. (2014) followed an analysing method derived from Forrest et al. (1988), which segmented the frication part into a few windows, and captured the chronological trend of the spectral moments of each fricative. Comparing English and

Mandarin fricatives in terms of the ability of the spectral moments to distinguish categories, it appeared that each moment performed differently in different languages. Spectral mean measurement could distinguish pairs of Mandarin fricatives which were next to each other in terms of place of articulation, while it could not distinguish between English /f/ and /θ/. In contrast, spectral variance measurement distinguished all places of articulation in English, while it could only distinguish two Mandarin fricative pairs: /f/-/ɕ/ and /f/-/ʂ/. Spectral skewness was able to distinguish all English fricatives and most of the Mandarin fricatives except between /f/ and /ɕ/. Spectral kurtosis could also distinguish all English fricatives, but only two pairs of Mandarin fricatives, including /f/-/s/ and /f/-/ɕ/.

Noticeably, Jongman et al. (2000) did not mention one important measurement of fricative, which was the centre of gravity (CoG) of fricatives, while Lee et al. (2014) mixed the term CoG with the first spectral moment, which was referred to as the spectral mean by both studies. CoG was defined as the average of frequency over the entire frequency domain, weighted by the power/intensity of its respective frequency (Boersma & Weenink, 2018), and was used in a number of studies on fricative acoustic properties (e.g. Gordon, Barthmaier, & Sands, 2002; Johnson, 1991; Nittrouer, 1986; Yu & Lee, 2014). In a study by Maniwa, Jongman, and Wade (2009), the first spectral moment analysed was CoG. As Jongman et al. (2000) pointed out, spectral mean aimed to reflect the average energy concentration, and the calculation was conducted following the detailed calculation introduced by Forrest et al. (1988). In their report, they stated that the spectral mean was an index for CoG of a spectrum. Despite the confusion regarding the usage of terms, it is safe to draw the conclusion that CoG and spectral mean were measurements serving the same purpose, which was to locate the energy centre of a spectrum.

2.1.2 Transition information measurements

Formant transition cues refer to acoustic information taking place in a time window centred on the boundary between two sounds, which contain information of both sounds. For fricatives, F2 is considered to contain meaningful information that indicates different places of articulation of fricatives. F2 is the frequency of the second formant, which is associated with the resonances shaped by the oral cavity posterior to

the constriction (Jongman et al., 2000), so it could serve as an indicator for the location where the constriction is formed in the mouth. F2 measurements at vowel onset is the frequency of the second formant at the first glottal pulse of a vowel succeeding a fricative. Locus equation is a measurement that links the vowel onset F2 frequency and its corresponding mid-vowel frequency, tracking the transition of F2 from a fricative to the middle of a vowel (Sussman et al., 1991; Sussman & Shore, 1996). The F2 measurement at the fricative-vowel boundary was found to be an indicator of fricatives' places of articulation (Wilde, 1993), while locus equation models the trajectory of F2 change linked to the transition from a fricative to a vowel (Sussman et al., 1991). Both Jongman et al. (2000) and Lee et al. (2014) reported measurements of F2 onset, while the latter did not report any locus equation analysis. For Mandarin fricatives, F2 onset could distinguish almost all fricatives, except between /s/ and /ʂ/. Similarly, for English fricatives, F2 onset could distinguish almost all fricatives, except between /s/ and /θ/. F2 onset appeared to be an efficient acoustic property that could contribute the same amount for fricative categorisation in English and Mandarin. Nevertheless, these two languages may weigh F2 onset as an acoustic cue differently. Considering that Mandarin learners of English tend to assimilate /θ/ to their /s/ category, it indicates that it could be because they rely on F2 onset more than English native listeners. In terms of locus equation, it appeared to be only able to distinguish /f/ from the other fricatives. Despite its limitation in distinguishing among places of articulation, the measure of locus equation could distinguish the two non-sibilants /f/ and /θ/, which formed a contrast that was difficult to distinguish even for native speakers (Forrest et al., 1988; Harris, 1958; Heinz & Stevens, 1961).

2.1.3 Frication amplitude

Frication amplitude may be evaluated in two ways: either by measuring the overall amplitude value of the frication independently, or by normalising frication amplitude measurements by relating them to the values of the following vowels of the fricative syllables. Both overall and relative measurements can be informative for fricative studies that recorded the production of individual syllables, as they are more likely to be produced consistently. In natural speech, on the other hand, the relative amplitude should be a more informative measurement, as each fricative token embedded in a

speech stream may vary greatly due to other linguistic factors of natural speech, such as the speech rhythm, stress pattern, and intonation.

Jongman et al. (2000) and Lee et al. (2014) both recorded only individual fricative syllables, but they included different measurements in their reports. Jongman et al. (2000) reported both overall and normalized measurements of fricative amplitude and duration, while Lee et al. (2014) only reported normalized amplitude and duration. As a result, we could only compare the normalized measurements of English and Mandarin fricatives.

For relative amplitude, these studies conducted two ways of measuring. One of the methods was the difference between the overall frication amplitude and the vowel amplitude at specific frequency ranges. In Jongman et al.'s study (2000), the vowel amplitude was measured at specific frequency region according to the preceding fricative: F3 amplitude was measured for sibilants, and F5 was measured for non-sibilants. This method was based on the results of Hendrick and Ohde's perceptual study (1993) on fricatives. They conducted an identification task, which revealed that manipulating the relative amplitude at F3 region could influence the identification result of the /s/-/ʃ/ pair, and manipulating at F5 region could influence the result of the /s/-/θ/ pair. This result was taken by Jongman et al.'s study (2000) as that relative amplitude of F3 region was for distinguishing sibilants, and relative amplitude of F5 region was for distinguishing non-sibilants. Their analysis showed that relative amplitude measurements could distinguish all English fricatives' places of articulation. In comparison, Lee et al.'s study (2014) calculated relative amplitude of each Mandarin fricative at F3, F4, and F5 regions, and it turned out that relative amplitude at different regions could distinguish different fricative pairs. At F3 region, it could distinguish /ʃ/ from /s/ and /f/; at F4 region, it could distinguish most fricative pairs except /f/-/s/ and /ʂ/-/ɕ/; and relative amplitude at F5 region was unable to distinguish any fricative pairs.

Another method to measure the relative amplitude between fricative and vowel was calculating the relative root-mean-square amplitude in dB, or the normalised rms amplitude. Both studies referenced the method introduced in a study by Behrens and Blumstein (1988b), which calculated the overall amplitude of the frication portion, and

the average amplitude of the three strongest consecutive glottal pulses of the vowel portion, and the difference between the two values was the relative amplitude. For the amplitude of vowels, Jongman et al. (2000) calculated the rms value of the three pulses instead of their average, and Lee et al. (2014) also followed them. In Jongman et al.'s study (2000), the relative rms amplitude was able to distinguish all four places of articulation of English fricatives, as each fricative's relative rms amplitude was significantly different from the others'. In addition, the difference was greater between sibilant group and non-sibilant group than between the fricatives within groups. In comparison, the relative rms amplitude was able to distinguish most of the Mandarin fricatives from each other, apart from two fricative pairs: /s/-/ɕ/ and /ɕ/-/ʂ/. Though this measurement appeared to be less efficient at distinguishing Mandarin fricatives than at English fricatives, Mandarin /f/ was distinct to this measurement.

Acoustically, relative amplitude had demonstrated its efficiency in distinguishing fricatives, especially English fricatives. However, perceptually, studies have found that relative amplitude as a cue contributed little to fricative identification, as changes in relative amplitude did not have an effect on perception when spectral properties and formant transitions were compatible; moreover, the relative amplitude and spectral properties of a fricative do not change independently in natural speech (Behrens & Blumstein, 1988b; Hendrick & Ohde, 1993). In other words, variations in relative amplitude would also entail changes in spectral properties of fricatives. Hendrick and Ohde (1993) admitted in their report that, despite discovering the significant effect of relative amplitude in the perception of place of articulation of fricatives, spectral property changes could be the actual predominant cue that had determined fricative identification. Whether relative amplitude is relevant in fricative categorisation was not yet clear. Its implication on fricative perception could not be discussed on isolation without inspecting the spectral changes taking place simultaneously. In addition, when comparing the acoustic studies by Jongman et al. (2000) and Lee et al. (2014), it was noticeable that they measured the relative amplitude differently: one chose a specific frequency range depending on if it was a sibilant, and one chose 3 specific frequency ranges for all the fricatives. An appropriate method measuring relative amplitude at specific frequency range appeared to be language specific. For a cross-language comparison, this way of measuring relative amplitude may not be suitable.

2.1.4 Frication duration

Similar to frication amplitude, frication duration could also be analysed with two measures: one was to measure the absolute duration of the frication alone, and another was to calculate the relative duration which was the ratio of frication duration over syllable/word duration. The absolute frication duration appeared to be able to discriminate sibilant group from non-sibilant group in English (Behrens & Blumstein, 1988b), and this result was replicated by Jongman et al.'s study (2000). The latter study also reported analyses of relative duration, and it performed better than absolute duration in terms of distinguishing fricatives: it could discriminate most fricative pairs in English, excluding /f/-/θ/. Lee et al.'s study (2014) of Mandarin fricatives only reported analyses of relative duration, and in the pair-wise comparisons only /f/ stood out as the shortest amongst all fricatives, when the other fricatives were not significantly different from each other. As /f/ is the only non-sibilant in Mandarin, relative duration appeared to perform the same duty in Mandarin as in English, which was distinguishing non-sibilants from sibilants. It was unclear how the function of relative duration may lead to the same or different dependence on this cue during fricative perception of English and Mandarin listeners. Notably, frication duration may vary depending on other features of speech, such as speech style and rhythm (e.g. syllable-timed or stress-timed), syllable structure, and intonation (Mok, 2009; Wang, Zhang, & Xu, 2018; Xu & Wang, 2009). Consequently, all these factors cannot be left out of a cross-linguistic comparisons of segment duration within connected speech; otherwise, comparing the different perceptual roles of frication duration across languages cannot lead to reliable conclusions.

In terms of the perceptual weighting of frication duration, studies were holding conflicting views. Jongman (1985) drew the conclusion that fricative duration, which he also referred to as fricatives' temporal properties, was not an important cue for identifying place of articulation, while spectral properties were playing a major role. His study discovered that shortening the fricatives by 50 ms in syllables did not significantly affect listeners' perceptual performance as their accuracy was still above chance level. Amongst all the English voiceless fricatives, the performance of identifying /θ/ was the most affected by the decrease in duration. Another study by Jongman (1989) on the perceptual effect of fricative duration challenged his previous

view on it, as the results revealed that a duration shorter than 40 ms could not provide sufficient information for an accurate identification of most places of articulation of fricatives except /ʃ/. In a later study on fricative acoustics across seven languages, frication duration was once again deemed as a "poor differentiator" of fricatives, as it appeared to be less useful for differentiating fricatives when compared to spectral properties such as CoG and spectral shape (Gordon et al., 2002, p. 166). Even /f/, which was found to be consistently shorter than sibilants in both Mandarin and English, could be longer than /ʃ/ in some languages. Lee et al. (2014) also suggested that frication duration may not serve as a dependable cue for distinguishing Mandarin fricatives. Overall, fricative duration does not seem to play an important role in fricative perception.

2.1.5 Aims of the present study

As the main research question was to investigate the cross-language differences in fricative cue processing and weighting, it was crucial to have a clear understanding of what cues were available. So far there was lack of studies of Cantonese fricative acoustics, and there was yet a study that analysed and compared fricative acoustics of English, Mandarin, and Cantonese. The present study aimed to investigate what acoustic cues, especially spectral cues, from natural speech stream were more statistically significant for fricatives of each language, thus potentially more reliable, for fricative perception. More importantly, this study attempted to reveal the differences in acoustics of the fricatives of Mandarin and Cantonese, which potentially had led to their different assimilation patterns of /θ/. The target phonemes were voiceless fricatives of Cantonese, English, and Mandarin. The plan was to record speech streams with target fricatives embedded within, and analyse the physical features of these targets. In order to analyse the physical features of the fricatives, the study adopted some of the measurements of Jongman et al. (2000) and Lee et al. (2014), including spectral property, temporal property measurements. The analysis of each target fricative would then be gathered to perform within- and cross-language comparisons. Following the classic acoustic analysis of the production materials, an attempt to create fricative identifier models using a machine-learning approach was presented. These models intended to imitate phonologically-naïve, monolingual "listeners", who are only exposed to the acoustic information of one fricative from

production data. It could not only be compared to the classic acoustic analysis, but also show insights to the different assimilation patterns of /θ/ of human listeners with different native languages.

The present study conducted an acoustic comparison of fricatives from the Cantonese, Mandarin and English, which was not done by other studies before. Instead of comparing results from various studies which may adopt different acoustic measurements, this study recorded, measured, and analysed the fricatives using the same method and in the same controlled environment. It allows a valid and direct comparison among acoustic features of all the target fricatives, generating more convincing conclusions.

Other than including Cantonese fricatives in the acoustic analysis, this study intended to make two adjustments to improve the method used in the previous studies (Jongman et al., 2000; Lee et al., 2014). The first adjustment was to obtain fricatives from near-natural speech streams. Compared to recording isolated syllables or words, acquiring fricatives produced in speech stream could maintain more natural acoustic cues closer to those from real life speech. As a result, these acoustic properties would better equip us for understanding L2 perception in a natural setting. The second adjustment was to acquire more fricative tokens from recordings through recruiting more individuals than the number of subjects recorded in Lee et al. (2014), which was only 6. Considering the noticeable individual differences in fricative production (Hughes & Halle, 1956; Yu, 2016), a larger sample size could minimize the effect of individual variances, and reveal phonologically distinct acoustic differences among fricative categories of different languages.

A few more adjustments were made in the present study in order to suit the research question of the project. One of the adjustments was to only include fricatives relevant to the main research question. Since the present study aimed to provide a foundation for a better understanding of the assimilation patterns of /θ/ of Cantonese and Mandarin-speaking English learners, the target fricatives included in this study were limited to fricatives produced in between the lips and the hard palate, including labiodental, dental, alveolar, and post-alveolar/retroflex. The Mandarin alveolo-palatal fricative /ɕ/ was excluded in this study, as there were studies debating over its role as

a phoneme in Mandarin. It directly preceding a low vowel was considered phonotactically illegal by some studies, and thus its phonemic position was still questioned (e.g. Duanmu, 2007; Lu, 2014). Despite the debate over /ɕ/ about if it was a phoneme in Mandarin, its longer constriction channel formed between the tongue and alveolo-palatal region lead to significantly different spectral cues (Lee et al., 2014), and it was not a viable candidate for /θ/ assimilation.

Another adjustment was limiting the vowel contexts of where the target fricatives occurred. Only the fricatives that occurred before unrounded open vowels / æ a ɑ / were used for analysis in the present study (vowel quality of these vowels may be different across languages, though they share the same IPA symbols; these differences are not a focus of the current study). For CV syllables with initial fricatives, unrounded open vowels were the only type of vowels that are phonotactically acceptable in all three languages. Meanwhile there are different phonotactic limitations regarding fricatives being followed by close vowels in English, Mandarin, and Cantonese. For instance, a front close vowel (e.g. /i/) following a fricative is allowed by English, while there are only words following /s/ but not /f/ in Cantonese, and it is completely not allowed to follow either /f/ or /s/ in Mandarin (Duanmu, 2007; Lai & Cheng, 1991). Analysing fricatives under only / æ a ɑ / contexts would allow convenience for comparisons among fricatives across languages.

The study included analysis of all of the spectral measurements, including spectral peak location, and all four spectral moments, and F2 onset frequency at fricative-vowel boundary. As discussed in section 2.1.1, the first spectral moment analysed in the current study was referred to as CoG instead of spectral mean. And since this study was aiming to provide acoustic information for enhance the understanding of perception, it would not include any absolute or relative amplitude and duration measurements due to their co-dependent status and/or their limited impact on perception (detailed reasoning provided in section 2.1.3 and 2.1.4). The measure of locus equation was also not included, as it has to come with a discussion about cross-language vowel variabilities, which was not a focus of this study. Since the present study analysed natural reading, and the fricative syllables occurred in connected speech rather than independently, the vowels within these syllables would contain transitional acoustics anticipating the following phonemes, affecting the acoustic

properties of the vowels, making it problematic to pluck out the coarticulatory effect in the middle of a vowel caused only by syllable-initial fricatives. In addition, in order to obtain more fricative tokens in natural speech, vowels were not strictly controlled in number and quality, and even if a vowel phoneme is shared by languages their qualities may differ. As a result, transitional information was only measured by F2 onset frequency in the classic acoustic analysis.

The fricative identifier models served as a discriminant analysis, which was performed to evaluate the acoustic similarities and dissimilarities among fricatives. It may also serve as a supplement to the classic acoustic analysis, as it was not limited by the same constraints discussed above. This machine-learning based technique was applied to find time-frequency functions that could recognise target acoustics from the spectrograms. This method was inspired by two-dimensional linear functions of sensors against time used in EEG studies (i.e., mTRF; Crosse, Di Liberto, Bednar and Lalor, 2016) in order to fit an EEG recording back to the original signal. These were used, for example, to map the EEG recording back to the amplitude envelope of the speech stimuli, or to analyse discrete events such as the onsets of words. The fit two-dimensional matrix of weights is a time-domain filter that can be convolved with the EEG recording to model what was originally heard, which in principle can also be used with speech recordings to model what acoustic cues are necessary for identifying a specific fricative. With this method going alongside the traditional acoustic analysis, the present study intended to provide reliable cross-linguistic comparisons for fricative acoustics of English, Mandarin, and Cantonese.

2.2 Method

2.2.1 Participants

This experiment recruited 21 standard Southern British English native speakers (10 males and 11 females), 16 Northern Mandarin native speakers (6 males and 10 females), and 13 Hong Kong Cantonese speakers (3 males and 10 females) to read the stories. The native English speakers that were recruited had no learning experience of either Mandarin or Cantonese. The Mandarin and Cantonese speaking participants did not start learning English as a module before they started primary school education, and they had been exposed to an English-speaking environment for less than 3 years.

2.2.2 Procedure

The recordings were conducted in a soundproof booth, with a Rode NT-1A microphone positioned 45 cm away from their mouths, and an RME Fireplace UC audio interface. All recordings were sampled at 44 kHz using Audacity installed on a PC.

Real words with syllable-initial Cantonese /f/ and /s/, English /f/, /s/, /θ/, /ʃ/, Mandarin /f/, /s/, and /ʂ/ attaching to open vowels (including /a/, /ɑ/, and /æ/) were imbedded in three short stories. Each fricative occurred 8 times in the stories. The participants were instructed to read the stories naturally as if they were telling a story to a friend, and were left alone in the booth so that they would be comfortable to read naturally. Their recording performance was monitored from outside the booth. Each participant read the story in their native language 3 times, from which 2 clearer recordings, providing 16 tokens of each fricative, were selected for further analysis. The study collected 208 tokens for each Cantonese fricative, 336 tokens for each English fricative, and 256 tokens for each Mandarin fricative.

2.2.3 Analysis

The recordings went through a Hann band filter to eliminate background noises that were lower than 75 Hz, and their scale intensity was modified to 67 dB. The target fricative syllables were annotated and extracted using Praat (Boersma & Weenink, 2018), starting from a point where the frication noise began and the voicing (if any) from the previous sound stopped, which was cross-checked with spectrogram and waveform provided by Praat. The transition point was marked at the beginning of harmonic structure, defined as the zero-crossing of the first glottal pulse that appears on the sound waveform, as shown in Figure 2-1. This point was treated as an end of the frication noise (i.e. fricative offset) and vowel onset in the following analysis, following Jongman et al. (2000). The fricative offset and the vowel onset may have a tiny time gap in reality, but it is usually small enough to ignore.

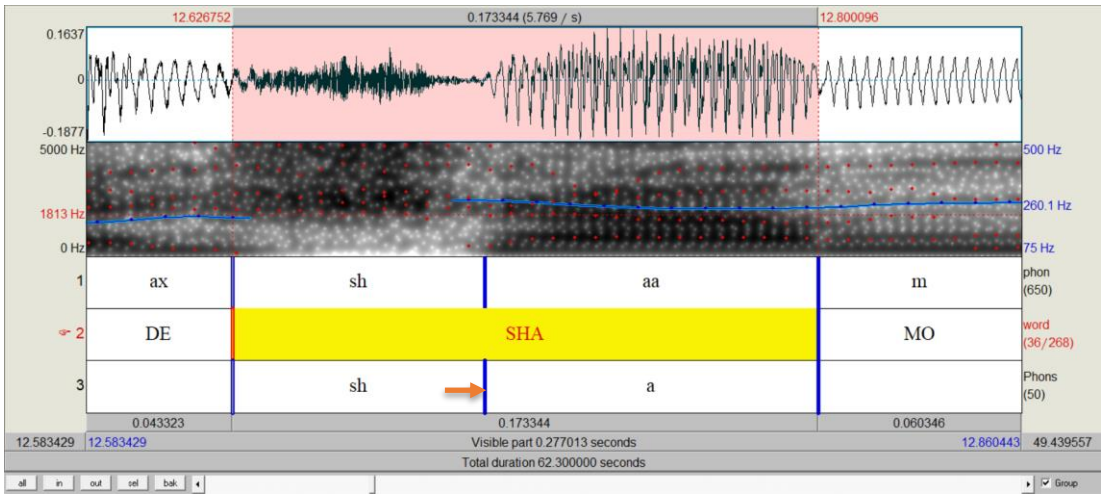


Figure 2-1 A screenshot of Praat annotation and segmentation of a recording of this study. The highlighted section is a segmented Mandarin syllable /ʂa/, marking the syllable onset, syllable final, and the boundary between the fricative and the vowel. The boundary/transition point between fricative and vowel is marked with an orange arrow, which is the point of zero crossing of the first harmonic resonance.

2.2.3.1 Spectral peak location

The spectral peak location of each fricative token was measured and analysed using Praat (Boersma & Weenink, 2018) and R. In Praat, a 40 ms Hamming window was applied in the middle of the fricative token in order to obtain a relatively high resolution on the frequency domain, following Jongman et al. (2000) and Lee et al. (2014). Then a fast Fourier transform (FFT) was conducted to derive a spectrum of the token. Subsequently, Praat output the calculated frequencies and their respective power level in dB/Hz of the spectrum. R was then used to locate the frequency with the highest power level within the spectrum output generated by Praat. One spectral peak was located for each fricative token.

2.2.3.2 Spectral moments

Spectral moments were calculated using MATLAB and Praat (Boersma & Weenink, 2018). Each fricative token was segmented into 3 windows: onset, middle, and offset; each segment was 40 ms long, and had a 40 ms Hamming window applied to it. This analysis skipped the forth window used in Jongman et al. (2000), which aimed to measure the transition from fricative to vowel. The reason for this was that, as McMurray and Jongman (2011) pointed out, window 3 and window 4 overlapped for 20 ms, and would thus violate the independence assumption of statistical tests. Moreover, considering the aim of spectral moment measurements which was to characterize the frication noise (Jongman et al., 2000), and that there were other

measurements, such as locus equations and F2 onsets, which focused only on the transition from fricative to vowel, the present study have made the decision to skip window 4 in the analysis.

As the source of the data was natural speech stream, not all of the tokens were longer than 40 ms for segmentation. In the end, from the Mandarin recordings, 215 tokens of /f/, 251 tokens of /s/, and 251 tokens of /ʂ/ were analysed; from the Cantonese recordings, 187 tokens of /f/, and 200 tokens of /s/ were analysed; from the English recordings, 332 tokens of /f/, 299 tokens of /θ/, and all the /s/ and /ʃ/ tokens (336 each) were analysed. The 4 spectral moments of each segment were calculated using the built-in functions of Praat (Boersma & Weenink, 2018).

2.2.3.3 F2 frequency at vowel onset

The F2 at vowel onset at the fricative-vowel boundary was measured and analysed using Praat. In Praat, the function *To Formant (burg)…* was used for a short-term spectral analysis of each vowel token following the fricatives extracted from recordings, which approximates the spectrum of each analysis frame by a number of formants (Boersma & Weenink, 2018). The analysis window length was set to be 25 ms starting at the fricative-vowel boundary (refer to Figure 2-1). After the approximation, the F2 values were extracted.

2.2.3.4 Fricative identifier model

For each recording, cochlear-scaled spectrograms were calculated. Each recording was passed through a gammatone filter bank, 36 ERB-scaled (Equivalent Rectangular Bandwidth) bands with centre frequencies from 63 to 19758 Hz. For each filter, the amplitude envelope was calculated using a Hilbert transform, was low-pass filtered at a 16 Hz cut-off frequency, and was down-sampled to 64 Hz.

The model fits were made using ridge regression that forced the model to use more data, in which case each channel × time point was a variable in the regression model. In the present application, as opposed to applying this model to EEG data when we fit information from multiple channels, the separate channels were spectral channels that were outputs of the gammatone filters. We fitted linear time × frequency functions that, when convolved with the spectrograms, modelled the event of the fricative

transition to the vowel. For example, we fitted /f/ models that produced a peak value at every point an /f/ fricative transitioned to a vowel, and separated models for the other fricatives. The time window was +/- 100 ms from the transition point (demonstrated in Figure 2-1), enabling this model to use both the preceding frication information and the transitions into vowels.

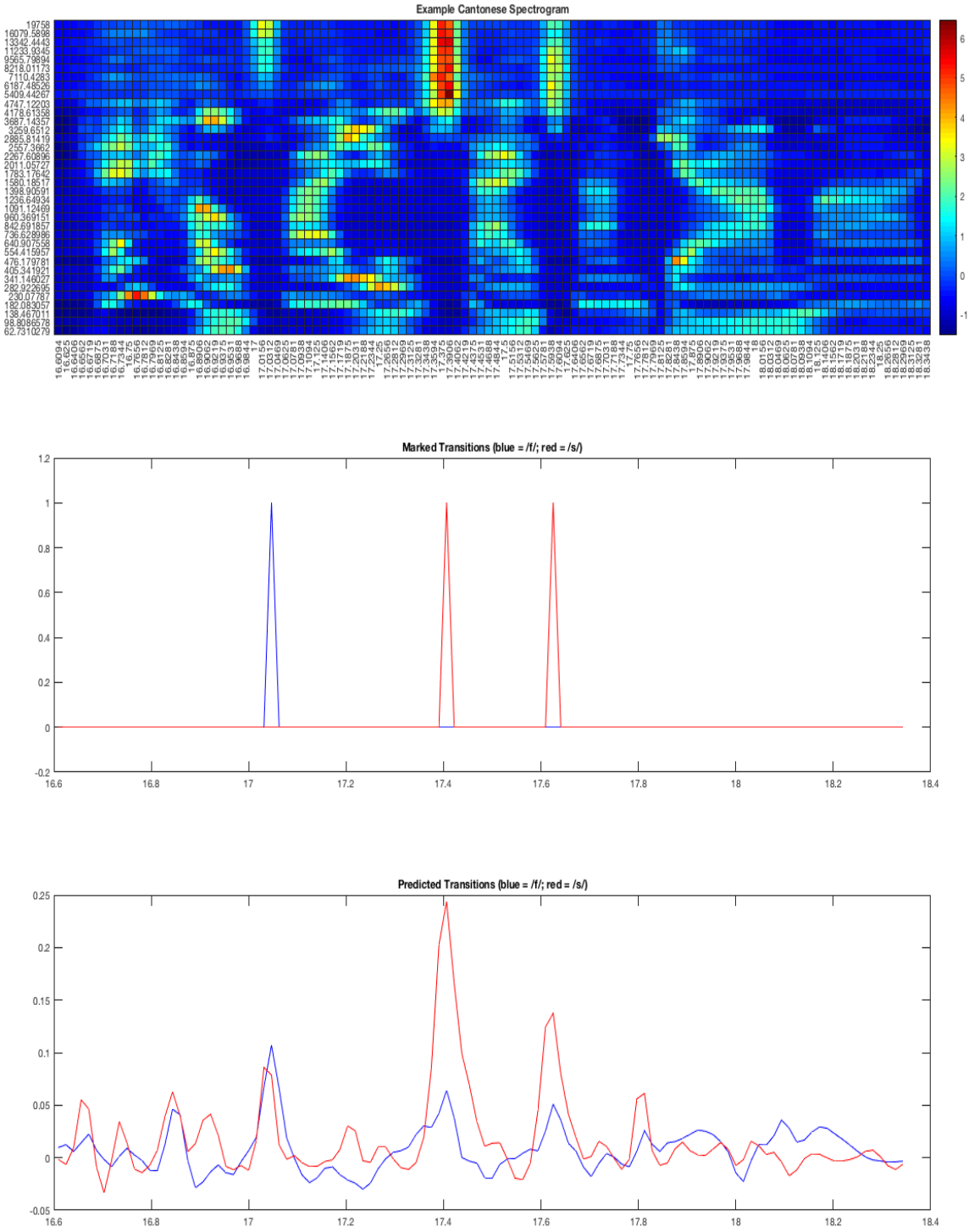


Figure 2-2 An example of how the models work. In this example, the Cantonese /f/ and /s/ models scan through a recording, and the models generates predictions of where the fricatives are. If the response reaches 1, it means that the model fits perfectly.

To fit the models, a leave-one-out cross-validation procedure was used. For example, for Cantonese /f/, each individual Cantonese speaker was left out, one at a time, from the training set with information from all other Cantonese speakers, so that the data left out could be used to evaluate the model. Each model was trained and built with only one fricative from a language, so that they were both fricative- and language-specific. The models were evaluated by applying them to the cochlear-scaled spectrograms, and assessing which model had a better response at the relevant fricative sections. That means, for example, the models for Cantonese /f/ and /s/ were applied to the Cantonese recordings, and for the points that had been previously identified as the /f/, we calculated whether the /f/ or /s/ model provided a better fit (i.e., producing a confusion matrix of the best-fit model for each fricative in the recording). If for a token, the /f/ model generated the highest score, it could be seen as if the token was identified as /f/. The models were also applied to all of the data to test how well the models fit both 'native' and 'non-native' fricatives. A percentage value was calculated for each model per token, which indicates how often each token was identified by a particular model.

2.3 Results

The statistical analyses were conducted in R. Unless stated otherwise, all measurement data from the classic acoustic analysis were fitted the peak location data of those phonemes into a linear mixed effect model with *lme4* package (version 1.1-23) fitted in R (Bates et al., 2015). Fricative token's source language (further referred to as *Language*), and fricative's phonemic category (further referred to as *Fricative*) were fixed factors, and subject number label (further referred to as *Subject*) was a random factor of the models. The full model formula in *lme4* style was e.g. peak.location ~ *Language* * *Fricative* + (1|*Subject*), to investigate both main effects of the fixed factors and their interactions.

Parameter estimates of the models were examined, and Likelihood Ratio Test (LRT) was conducted to investigate statistical significance of the factors and interaction following the method introduced by Winter (2013), with the *anova* function embedded in *lme4*. The method compares the full model against a reduced model without the

effects in question, and a fixed effect is considered significant if the difference between the likelihood of these two models is significant ($p < .05$).

According to the aim of the present study, which was investigating cross-linguistic differences in fricative cues and their perceptual weight, the interaction between *Language* and *Fricative* was the main interest of this analysis. For the purpose of this study, and following the suggestion from Chen, Xu, Tu, Wang and Niu (2018), a Tukey post hoc test was conducted on the interaction between *Language* and *Fricative* for each measure, with *emmeans* package (version 1.5.1, Lenth, 2020), for within and across languages pairwise fricative comparisons.

2.3.1 Spectral peak location

A descriptive summary of peak location data is reported by Table 2-1 and Figure 2-3. Amongst all the target fricatives, /s/s appeared to have the highest mean/median peak location, while /ʃ/ and /ʂ/ had the lowest mean/median. /f/s and /θ/ generally had higher variability compared to the other fricative categories. Cantonese and Mandarin /f/s had similar spectral peak locations while significantly differed from English /f/. Meanwhile, Mandarin /s/ had its peak at a significantly higher frequency when compared to English /s/, when Cantonese /s/'s peak fell in between the peaks of Mandarin /s/ and English /s/.

Table 2-1 Mean (M) and Standard deviation (SD) values (in Hz) of spectral peak location of all the target fricatives.

Language		**Fricative**				
		f	θ	s	ʃ	ʂ
Cantonese	M	4370		7598		
	SD	5273		2737		
English	M	5890	5250	7013	3358	
	SD	4851	4931	1864	867	
Mandarin	M	4374		8507		3064
	SD	5013		2579		1053

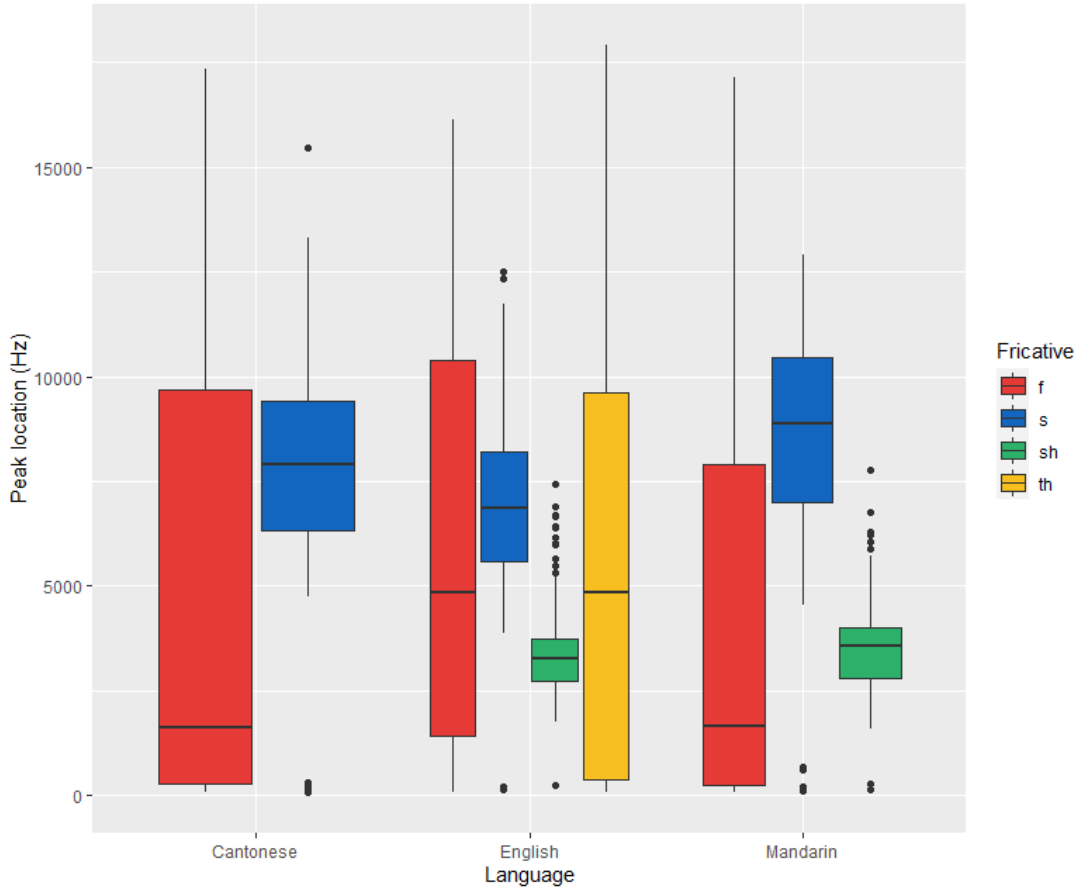


Figure 2-3 Peak locations in Hz of all target fricative tokens, grouped by native languages and fricative categories. Fricative label 'f' stands for /f/, 's' stands for /s/, 'sh' stands for /ʃ/ and /ʂ/ in English and Mandarin respectively, and 'th' stands for /θ/. C is Cantonese, E is English, and M is Mandarin.

The mixed-effect model analysis revealed a significant main effect of *Fricative* ($\chi^2(3) = 450.69$, $p < .001$), and a significant interaction between *Language* and *Fricative* ($\chi^2(3) = 54.99$, $p < .001$). To further investigate the interaction and whether the "same" fricative is different across languages, a subsequent Tukey's post hoc test was conducted on the interaction. The test results are shown in Table 2-2 and Table 2-3.

Within each language, the peak location measure could distinguish most of the fricative pairs. The fricative pairs that were not differentiated by peak location were Mandarin /f/ and /ʂ/, and English /f/ and /θ/.

For across language comparisons, it appeared that both Cantonese /f/ and Mandarin /f/ were significantly different from English /f/, and Mandarin /s/ was significantly different from English /s/. These differences are also demonstrated by Table 2-1 and Figure 2-3: the means and medians of Cantonese and Mandarin /f/s were lower than English /f/; while Mandarin /s/ had the highest mean and median amongst all the /s/s.

Mandarin /ʂ/ and English /ʃ/ were not significantly different in terms of peak location. The post hoc test also compared across fricative categories to explore potential effect of spectral peak location for the assimilation patterns of English /θ/. The results shows that both Mandarin and Cantonese /f/ were similar to /θ/, while neither Mandarin nor Cantonese /s/ were similar to /θ/.

Table 2-2 The output of the post hoc test with within-language fricative comparisons. In the 'Contrast' column, 'C' stands for Cantonese, 'M' is Mandarin, and 'E' is English; 'f' represents the fricative category /f/, and 's' represents the fricative category /s/. 'sh' stands for /ʃ/ and /ʂ/ in English and Mandarin respectively, and 'th' stands for /θ/.

Contrast	**Estimate**	**SE**	***t*-ratio**	***p*-value**
C f – C s	-3228.2	351	-9.203	*<.0001
M f – M s	-4133.4	316	-13.073	*<.0001
M f – M sh	913.0	316	2.888	0.0921
M s – M sh	5046.4	316	15.961	*<.0001
E f – E s	-1123.1	276	-4.070	*0.0016
E f – E th	634.0	276	2.294	0.3462
E f – E sh	2531.9	276	9.174	*<.0001
E s – E sh	3655.0	276	13.244	*<.0001
E s – E th	1757.1	276	6.357	*<.0001
E sh – E th	-1897.9	276	-6.866	*<.0001

Table 2-3. The output of the post hoc test with cross-language fricative comparisons. In the 'Contrast' column, 'C' stands for Cantonese, 'M' is Mandarin, and 'E' is English; 'f' represents the fricative category /f/, and 's' represents the fricative category /s/.

Contrast	**Estimate**	**SE**	***t*-ratio**	***p*-value**
C f – E f	-1369.9	321	-4.265	*0.0007
C f – M f	19.4	336	0.058	1.0000
E f – M f	1389.4	300	4.625	*0.0001
C s – E s	735.2	321	2.289	0.3491
C s – M s	-885.8	336	-2.636	0.1721
E s – M s	-1620.9	300	-5.396	*<.0001
C f – E th	-735.9	321	-2.290	0.3487
C s– E th	2492.3	321	7.754	*<.0001
M f– E th	-755.4	301	-2.512	0.2263
M s– E th	3378.1	301	11.236	*<.0001
E sh– M sh	-229.5	300	-0.764	0.9978

2.3.2 Spectral moments

Figure 2-4 demonstrated four spectral moments of the three languages, measured at three window locations: onset, middle, and offset. Descriptive data including mean and standard deviation is reported by Table 2-4. The four moments were analysed in separate models, each used *Language* and *Fricative* as fixed factors, and *Subject* as a random factor. The data were averaged over the factor *Window*, because analyses performed on the individual window did not differ across languages ($p > .05$).

Table 2-4 Mean (M) and Standard deviation (SD) values of spectral moments of all the target fricatives.

Measure	**Language**		**Fricative**				
			f	θ	s	ʃ	ʂ
CoG (Hz)	Cantonese	M	7035.17		8041.26		
		SD	2949.79		1476.12		
	English	M	7261.28	7001.03	7321.37	4140.55	
		SD	2492.63	2865.74	1510.00	849.35	
	Mandarin	M	7139.72		8593.69		4370.24
		SD	2681.63		1491.45		743.45
Variance (Hz)	Cantonese	M	4535.24		2346.51		
		SD	982.09		694.80		
	English	M	4064.10	3885.29	2034.44	1739.86	
		SD	906.77	952.39	540.48	393.51	
	Mandarin	M	4369.15		2467.03		1902.81
		SD	997.01		703.68		452.33
Skewness	Cantonese	M	0.48		0.11		
		SD	1.10		0.86		
	English	M	0.25	0.43	0.57	1.87	
		SD	1.03	1.49	0.92	0.97	
	Mandarin	M	0.48		-0.10		1.78
		SD	0.95		0.63		0.74
Kurtosis	Cantonese	M	0.68		2.83		
		SD	3.43		2.90		
	English	M	1.10	2.45	3.58	6.77	
		SD	6.96	33.48	4.23	6.98	
	Mandarin	M	0.70		2.04		6.17
		SD	3.37		2.30		5.89

2.3.2.1 CoG

The analysis revealed significance of both fixed factors: for *Language*, $\chi^2(2) = 105.81$, $p < .001$; for *Fricative*, $\chi^2(3) = 2838.20$, $p < .001$. The analysis also revealed a significant interaction between *Language* and *Fricative* ($\chi^2(3) = 120.51$, $p < .001$).

Since the present study was interested in potential language differences, post hoc tests were conducted to investigate the significant interaction between *Language* and *Fricative*, which means that the measures were averaged across window locations, and their results are shown in Table 2-5 and Table 2-6. When comparing fricatives within languages, it appears that the CoG measurement was sufficient in distinguishing all fricative categories in Cantonese and Mandarin, while it failed to distinguish the English fricative pairs /f/-/s/ and /f/-/θ/. When comparing fricative categories across languages, it appears that the CoG measurement was not able to distinguish the /f/ categories of different languages, while it could distinguish all the /s/ categories. Table 2-6 also showed comparisons between /θ/ and the target categories for its assimilation in Cantonese and Mandarin: neither of the /f/ categories of Cantonese and Mandarin was significantly different from /θ/, while both of the /s/ categories were different from /θ/.

Table 2-5 Post hoc test results including comparisons of CoG measurements within languages. In the 'Contrast' column, 'C' stands for Cantonese, 'M' is Mandarin, and 'E' is English. Fricative label 'f' stands for /f/, 's' stands for /s/, 'sh' stands for /ʃ/ and /ʂ/ in English and Mandarin respectively, and 'th' stands for /θ/.

Contrast	**Estimate**	**SE**	***z*-ratio**	***p*-value**
C f – C s	-998.69	112.1	-8.911	*<.0001
M f – M s	-1468.89	102.4	-14.346	*<.0001
M f – M sh	2749.04	102.4	26.846	*<.0001
M s – M sh	4217.93	98.3	42.891	*<.0001
E f – E s	-60.01	85.2	-0.704	0.9999
E f– E th	268.89	87.9	3.060	0.0919
E f– E sh	3121.43	85.2	36.618	*<.0001
E s– E sh	3181.44	85.0	37.435	*<.0001
E s– E th	328.90	87.6	3.754	*0.0095
E sh– E th	-2852.54	87.6	-32.560	*<.0001

Table 2-6 Post hoc test results including comparisons of CoG measurements across languages. In the 'Contrast' column, 'C' stands for Cantonese, 'M' is Mandarin, and 'E' is English. Fricative label 'f' stands for /f/, 's' stands for /s/, 'sh' stands for /ʃ/ and /ʂ/ in English and Mandarin respectively, and 'th' stands for /θ/.

Contrast	**Estimate**	**SE**	***z*-ratio**	***p*-value**
C f – E f	-67.81	102.9	-0.562	1.0000
C f – M f	-50.98	110.9	-0.460	1.0000
E f – M f	6.83	97.9	0.070	1.0000
C s – E s	880.87	100.5	8.761	*<.0001
C s – M s	-521.18	105.2	-4.954	*<.0001
E s – M s	-1402.05	93.4	-15.017	*<.0001
C f – E th	211.08	104.9	2.012	0.6853
C s– E th	1209.77	102.9	11.762	*<.0001
M f– E th	262.06	99.9	2.622	0.2670
M s– E th	1730.95	95.8	18.072	*<.0001
E sh– M sh	-365.56	93.4	-3.915	*0.0051

2.3.2.2 Spectral variance

The analysis revealed significant main effects of *Language* ($\chi^2(2) = 284.70$, $p < .001$) and *Fricative* ($\chi^2(3) = 7898.70$, $p < .001$). There were also two significant interactions between factors: *Language* and *Fricative* ($\chi^2(3) = 46.44$, $p < .001$).

The post hoc tests further investigated the interaction between *Language* and *Fricative*, and the results are shown in Table 2-7 and Table 2-8. It appears that the measurements of spectral variance could distinguish all fricative categories within the three languages, and most fricative categories across languages, except Cantonese /s/ and Mandarin /s/ which appeared to have similar spectral variances.

Table 2-7 Post hoc test results including comparisons of spectral variance within languages. In the 'Contrast' column, 'C' stands for Cantonese, 'M' is Mandarin, and 'E' is English. Fricative label 'f' stands for /f/, 's' stands for /s/, 'sh' stands for /ʃ/ and /ʂ/ in English and Mandarin respectively, and 'th' stands for /θ/.

Contrast	**Estimate**	**SE**	***z*-ratio**	***p*-value**
C f – C s	2190	43.1	50.797	*<.0001
M f – M s	1901	39.4	48.263	*<.0001
M f – M sh	2466	39.4	62.587	*<.0001
M s – M sh	564	37.8	14.919	*<.0001
E f – E s	2030	32.8	61.891	*<.0001
E f– E th	182	33.8	5.379	*<.0001
E f– E sh	2325	32.8	70.892	*<.0001
E s– E sh	295	32.7	9.029	*<.0001
E s– E th	-1848	33.7	-54.828	*<.0001
E sh– E th	-2143	33.7	-63.586	*<.0001

Table 2-8 Post hoc test results including comparisons of spectral variance across languages. In the 'Contrast' column, 'C' stands for Cantonese, 'M' is Mandarin, and 'E' is English. Fricative label 'f' stands for /f/, 's' stands for /s/, 'sh' stands for /ʃ/ and /ʂ/ in English and Mandarin respectively, and 'th' stands for /θ/.

Contrast	**Estimate**	**SE**	***z*-ratio**	***p*-value**
C f – E f	496	39.5	12.555	*<.0001
C f – M f	180	42.7	4.212	*0.0015
E f – M f	-316	37.6	-8.413	*<.0001
C s – E s	336	38.6	8.693	*<.0001
C s – M s	-109	40.5	-2.702	0.2251
E s – M s	-445	35.9	-12.407	*<.0001
C f – E th	678	40.3	16.825	*<.0001
C s– E th	-1512	39.5	-38.284	*<.0001
M f– E th	498	38.4	12.977	*<.0001
M s– E th	-1403	36.8	-38.128	*<.0001
E sh– M sh	-176	35.9	-4.900	*0.0001

2.3.2.3 Spectral skewness

There were significant main effects of the fixed factors *Language* ($\chi^2(2) = 60.32$, $p < .001$), and *Fricative* ($\chi^2(3) = 2389.60$, $p < .001$). There were also significant interactions including the interaction between *Language* and *Fricative* ($\chi^2(3) = 198.53$, $p < .001$).

Table 2-9 Post hoc test results including comparisons of skewness measurements within languages. In the 'Contrast' column, 'C' stands for Cantonese, 'M' is Mandarin, and 'E' is English. Fricative label 'f' stands for /f/, 's' stands for /s/, 'sh' stands for /ʃ/ and /ʂ/ in English and Mandarin respectively, and 'th' stands for /θ/.

Contrast	**Estimate**	**SE**	***z*-ratio**	***p*-value**
C f – C s	0.37542	0.0574	6.538	*<.0001
M f – M s	0.58419	0.0525	11.134	*<.0001
M f – M sh	-1.30052	0.0525	-24.786	*<.0001
M s – M sh	-1.88471	0.0504	-37.402	*<.0001
E f – E s	-0.32311	0.0437	-7.397	*<.0001
E f– E th	-0.18786	0.0450	-4.173	*0.0018
E f– E sh	-1.62319	0.0437	-37.161	*<.0001
E s– E sh	-1.30008	0.0435	-29.854	*<.0001
E s– E th	0.13526	0.0449	3.013	0.1046
E sh– E th	1.43534	0.0449	31.974	*<.0001

Table 2-10 Post hoc test results including comparisons of skewness measurements across languages. In the 'Contrast' column, 'C' stands for Cantonese, 'M' is Mandarin, and 'E' is English. Fricative label 'f' stands for /f/, 's' stands for /s/, 'sh' stands for /ʃ/ and /ʂ/ in English and Mandarin respectively, and 'th' stands for /θ/.

Contrast	**Estimate**	**SE**	***z*-ratio**	***p*-value**
C f – E f	0.17604	0.0526	3.345	*0.0393
C f – M f	-0.01815	0.0568	-0.320	1.0000
E f – M f	-0.19419	0.0501	-3.876	*0.0060
C s – E s	-0.52249	0.0514	-10.158	*<.0001
C s – M s	0.19062	0.0539	3.537	*0.0207
E s – M s	0.71311	0.0478	14.926	*<.0001
C f – E th	-0.01182	0.0537	-0.220	1.0000
C s– E th	-0.38723	0.0526	-7.359	*<.0001
M f– E th	0.00634	0.0511	0.124	1.0000
M s– E th	-0.57785	0.0490	-11.789	*<.0001
E sh– M sh	0.12848	0.0478	2.689	0.2314

The post hoc tests investigating the interaction between *Language* and *Fricative* had shown details reported in Table 2-9 and Table 2-10. When comparing fricative categories within languages, the measurements of skewness averaged across window locations could distinguish all the fricatives of Cantonese and Mandarin, and most of the fricatives of English, except the fricative pair /s/ and /θ/. When comparing fricative categories across languages, skewness appeared to be able to distinguish all the /s/

categories, but not the /f/ categories as Cantonese /f/ and Mandarin /f/ were similar. When comparing the target assimilation categories of Cantonese and Mandarin to /θ/, we could see that both Cantonese /f/ and Mandarin /f/ shared similar skewness as /θ/, and both /s/s were significantly different from /θ/. In addition, Mandarin /ʂ/ and English /ʃ/ were not significantly different in terms of skewness measurements.

2.3.2.4 Spectral kurtosis

The calculated model revealed a significant main effect of *Fricative* ($\chi^2(3) = 184.12$, $p < .001$). There was a near significant main effect of *Language* ($\chi^2(2) = 5.94$, $p = .05$), but no significant interaction between *Language* and *Fricative* ($\chi^2(3) = 1.94$, $p = .58$).

The results of the post hoc test investigating the effect of *Language* and *Fricative* interaction are shown by Table 2-11 and 2-12. For within-language pairwise comparisons, kurtosis (averaged across window) could not distinguish both Mandarin and Cantonese /f/-/s/, and English /f/-/θ/ and /s/-/θ/. Different from other spectral moments, kurtosis could not distinguish any cross-language /f/ and /s/ comparisons. Moreover, kurtosis also could not distinguish Cantonese and Mandarin /f s/ from English /θ/.

Table 2-11 Post hoc test results including comparisons of kurtosis measurements within languages. In the 'Contrast' column, 'C' stands for Cantonese, 'M' is Mandarin, and 'E' is English. Fricative label 'f' stands for /f/, 's' stands for /s/, 'sh' stands for /ʃ/ and /ʂ/ in English and Mandarin respectively, and 'th' stands for /θ/.

Contrast	**Estimate**	**SE**	***z*-ratio**	***p*-value**
C f – C s	-2.1415	0.748	-2.864	0.0977
M f – M s	-1.3385	0.683	-1.960	0.5718
M f – M sh	-5.4749	0.683	-8.016	*<.0001
M s – M sh	-4.1364	0.656	-6.305	*<.0001
E f – E s	-2.4728	0.569	-4.348	*0.0005
E f– E th	-1.3420	0.586	-2.290	0.3482
E f– E sh	-5.6650	0.569	-9.961	*<.0001
E s– E sh	-3.1922	0.567	-5.630	*<.0001
E s– E th	1.1308	0.584	1.935	0.5891
E sh– E th	4.3230	0.584	7.398	*<.0001

Table 2-12 Post hoc test results including comparisons of kurtosis measurements across languages. In the 'Contrast' column, 'C' stands for Cantonese, 'M' is Mandarin, and 'E' is English. Fricative label 'f' stands for /f/, 's' stands for /s/, 'sh' stands for /ʃ/ and /ʂ/ in English and Mandarin respectively, and 'th' stands for /θ/.

Contrast	**Estimate**	**SE**	***z*-ratio**	***p*-value**
C f – E f	-0.4206	0.673	-0.624	0.9995
C f – M f	-0.0126	0.736	-0.017	1.0000
E f – M f	0.4079	0.644	0.633	0.9994
C s – E s	-0.7519	0.658	-1.143	0.9676
C s – M s	0.7904	0.698	1.133	0.9693
E s – M s	1.5423	0.614	2.513	0.2258
C f – E th	-1.7626	0.687	-2.566	0.2011
C s– E th	0.3789	0.673	0.563	0.9998
M f– E th	-1.7500	0.658	-2.660	0.1625
M s– E th	-0.4114	0.630	-0.653	0.9993
E sh– M sh	0.5981	0.614	0.974	0.9882

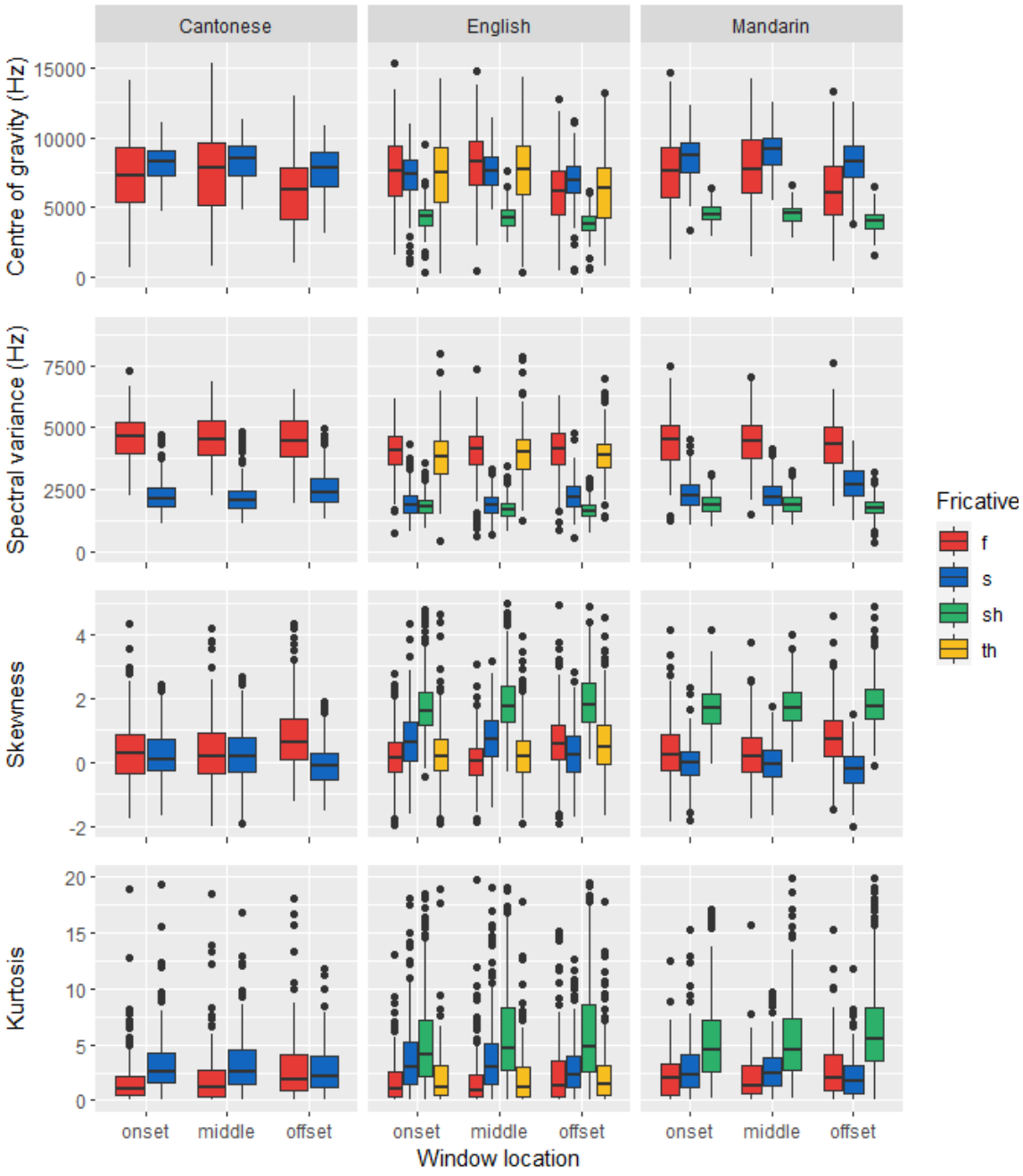


Figure 2-4 Boxplots consisting of measurements of four spectral moments, grouped by language, fricative category, and window location. Fricative label 'f' stands for /f/, 's' stands for /s/, 'sh' stands for /ʃ/ and /ʂ/, and 'th' stands for /θ/.

2.3.3 F2 frequency at vowel onset

A descriptive summary of F2 onset frequency data is reported by

Table *2-13* and Figure 2-5. F2 frequency appeared to increase as the place of articulation of the fricative moves further away from lips, and this trend was consistent across languages.

Table 2-13 Mean (M) and Standard deviation (SD) values (in Hz) of F2 frequency at vowel onset of all the target fricatives.

Language		**Fricative**				
		f	θ	s	ʃ	ʂ
Cantonese	M	1819.96		1949.35		
	SD	336.58		311.67		
English	M	1826.76	1920.58	1997.01	2225.81	
	SD	403.70	401.43	442.26	351.36	
Mandarin	M	1830.00		1978.45		2136.85
	SD	368.07		311.60		358.99

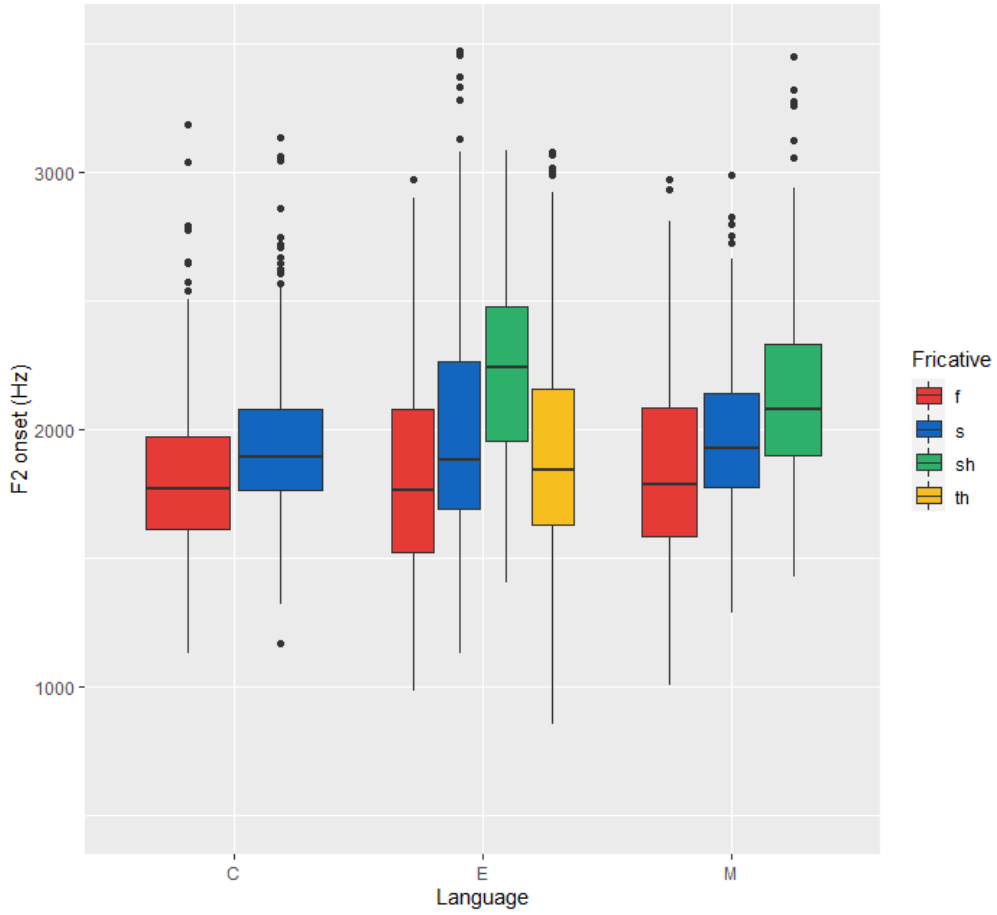


Figure 2-5 F2 frequency at vowel onset in Hz of all target fricative tokens, grouped by native languages and fricative categories. Fricative label 'f' stands for /f/, 's' stands for /s/, 'sh' stands for /ʃ/ and /ʂ/ in English and Mandarin respectively, and 'th' stands for /θ/. C is Cantonese, E is English, and M is Mandarin.

A significant main effect of *Fricative* ($\chi^2(3) = 292.87$, $p < .001$) and a non-significant main effect of *Language* ($\chi^2(2) = 3.04$, $p = .22$) was revealed, which confirmed the observation made based on Figure 2-5 and

Table 2-13 Mean (M) and Standard deviation (SD) values (in Hz) of F2 frequency at vowel onset of all the target fricatives.. There was not a significant interaction between the factors ($\chi^2(3) = 5.52$, $p = .14$). Tukey's post hoc test was conducted on the interaction. The test results are shown in Table 2-14 and Table 2-15. Within each language, the F2 onset measure could distinguish most of the fricative pairs. The only fricative pair that was not differentiated by F2 onset was English /s/ and /θ/. On the other hand, cross-language comparisons revealed no significant fricative contrasts.

Table 2-14 The output of the post hoc test with within-language fricative comparisons. In the 'Contrast' column, 'C' stands for Cantonese, 'M' is Mandarin, and 'E' is English; 'f' represents the fricative category /f/, and 's' represents the fricative category /s/.

Contrast	**Estimate**	**SE**	***t*-ratio**	***p*-value**
C f – C s	-129.395	36.2	-3.574	*0.0108
M f – M s	-148.458	32.6	-4.549	*0.0002
M f – M sh	-306.857	32.6	-9.402	*<.0001
M s – M sh	-158.399	32.6	-4.854	*<.0001
E f – E s	-170.255	28.5	-5.977	*<.0001
E f – E th	-93.821	28.5	-3.293	*0.0279
E f – E sh	-399.049	28.5	-14.008	*<.0001
E s – E sh	-228.794	28.5	-8.032	*<.0001
E s – E th	76.434	28.5	2.683	0.1543
E sh – E th	305.228	28.5	10.715	*<.0001

Table 2-15 The output of the post hoc test with cross-language fricative comparisons. In the 'Contrast' column, 'C' stands for Cantonese, 'M' is Mandarin, and 'E' is English; 'f' represents the fricative category /f/, and 's' represents the fricative category /s/.

Contrast	**Estimate**	**SE**	***t*-ratio**	***p*-value**
C f – E f	-0.103	33.1	-0.003	1.0000
C f – M f	-11.345	34.7	-0.327	1.0000
E f – M f	-11.243	31.0	-0.363	1.0000
C s – E s	-40.963	33.1	-1.238	0.9482
C s – M s	-30.408	34.7	-0.877	0.9941
E s – M s	10.555	31.0	0.341	1.0000
C f – E th	-93.924	33.1	-2.838	0.1051
C s– E th	35.471	33.1	1.072	0.9782
M f– E th	-82.578	31.0	-2.667	0.1604
M s– E th	65.879	31.0	2.127	0.4546
E sh– M sh	80.950	31.0	2.614	0.1811

2.3.4 Summary of acoustic measures

A summary of the acoustic measures is reported by Table 2-16. For Cantonese, most spectral measures could distinguish its fricatives, except kurtosis. For Mandarin, CoG, spectral variance, spectral skewness, and F2 onset were sufficient on their own as they could distinguish all the fricative pairs. For English, only spectral variance could discriminate all 6 of the English fricatives pairs, while peak location, skewness, and F2 onset differentiated 5 English fricative pairs. Interestingly, English /f/-/θ/ and /s/-/θ/ contrasts were two contrasts that were undistinguishable by the most measures: /f/-/θ/ appeared to have similar peak location, CoG, and kurtosis, while /s/-/θ/ appeared to have similar skewness, F2 onset, and kurtosis.

Table 2-16 Summary of significant fricative contrasts in the acoustic measures. C = Cantonese, E = English, M = Mandarin.

Language	**Fricatives**	**Peak location**	**CoG**	**Variance**	**Skewness**	**Kurtosis**	**F2 onset**
	Within-language comparisons						
C	/f/ - /s/	*	*	*	*		*
M	/f/ - /s/	*	*	*	*		*
	/f/ - /ʂ/		*	*	*	*	*
	/s/ - /ʂ/	*	*	*	*	*	*
E	/f/ - /s/	*		*	*	*	*
	/f/ - /θ/			*	*		*
	/f/ - /ʃ/	*	*	*	*	*	*
	/s/ - /ʃ/	*	*	*	*	*	*
	/s/ - /θ/	*	*	*			
	/θ/ - /ʃ/	*	*	*	*	*	*
	Cross-language comparisons						
C – M	/f/ - /f/			*			
C – E		*		*	*		
E – M		*		*	*		
C – M	/s/ - /s/		*		*		
C – E			*	*	*		
E – M		*	*	*	*		
C – E	/f/ - /θ/			*			
M – E				*			
C – E	/s/ - /θ/	*	*	*	*		
M – E		*	*	*	*		

Cross-language /f/ and /s/ comparisons revealed that they were different from each other in at least one measure. For /f/, both Cantonese and Mandarin /f/ were different from each other only in spectral variance, while being different from English /f/ in 3 measures. For /s/, Cantonese and English differed in 3 measures, while Mandarin and English differed in all measures except kurtosis.

When comparing Cantonese and Mandarin /f s/ to English /θ/, the measures revealed that only spectral skewness could distinguish both /f/s from /θ/. On the other hand, both Cantonese and Mandarin /s/ differed from /θ/ in all measures except kurtosis.

2.3.5 Fricative identifier model

Table 2-17 and Table 2-18 display the identification percentage of the models, in which the percentage values mean that when detecting a recording, how frequent a model demonstrates a stronger response than the other models (as shown by 'predicted transitions' in Figure 2-2) for the marked sections. In other words, for 'native' tokens, the percentage is how often a model detects the fricatives that it was trained to detect; and for 'non-native' tokens, the percentage indicates how similar a type of tokens were to the fricatives that the model was trained with.

When applying the models to detect 'native' fricatives, most models could detect 89-99% of the target segments correctly. The only model that appeared to be problematic is the English /θ/ model, since it only spot 19% of the marked /θ/ in the recordings, while 80% of /θ/ was identified by English /f/ model. When applying the models to detect 'non-native' fricatives, some models, such as the models of Cantonese /s/, Mandarin /f/, and English /f/, could detect more than 87% of the fricatives of the same phoneme label from the other languages. The models of Cantonese /f/ and Mandarin /s/ could detect more than 60% of other languages' /f/ and /s/ respectively. Interestingly, the English /s/ model could detect only around 40% of the Mandarin /s/s, which was detected more often by the English /f/ model. In addition, the models demonstrated assimilation patterns of the 'non-native' fricatives based on fricative acoustics alone. It appeared that the Mandarin /f/ model provided a better fit than other Mandarin fricative models for 87% of the English /θ/ tokens, and the Cantonese /f/ model fit 63% of the /θ/ tokens. Both the Cantonese and Mandarin /f/ models identified the majority of the English /θ/ tokens.

Table 2-17 Identification results of the recogniser models for 'native' fricatives. The y-axis is the fricatives information input, the x-axis is the models. C = Cantonese, M = Mandarin, E = English.

	C-f	C-s	M-f	M-s	M-ʂ	E-f	E-s	E-ʃ	E-θ
C-/f/	**89%**	11%							
C-/s/	1%	**99%**							
M-/f/			**95%**	4%	1%				
M-/s/			8%	**90%**	3%				
M-/ʂ/			1%	0%	**99%**				
E-/f/						**98%**	0%	0%	2%
E-/s/						10%	**89%**	0%	0%
E-/ʃ/						3%	4%	**93%**	0%
E-/θ/						80%	1%	0%	**19%**

Table 2-18 Identification results of the recogniser models for 'non-native' fricatives. The y-axis is the fricatives information input, the x-axis is the models. C = Cantonese, M = Mandarin, E = English.

	C-f	C-s	M-f	M-s	M-ʂ	E-f	E-s	E-ʃ	E-θ
C-/f/			**92%**	3%	4%	**85%**	2%	0%	13%
C-/s/			7%	**78%**	15%	26%	**73%**	0%	1%
M-/f/	**75%**	25%				**92%**	1%	0%	7%
M-/s/	5%	**95%**				55%	**44%**	0%	1%
M-/ʂ/	9%	**91%**				11%	13%	**74%**	2%
E-/f/	**62%**	38%	**88%**	8%	3%				
E-/s/	0%	**100%**	5%	**64%**	31%				
E-/ʃ/	10%	**90%**	3%	0%	**96%**				
E-/θ/	**63%**	37%	**87%**	11%	2%				

2.4 Discussion

This study provided acoustic measurements of fricatives extracted from Cantonese, Mandarin, and English native speakers' speech, and conducted within and across language comparisons. Within-language comparisons revealed how efficient the acoustic measurements were to distinguish all the fricatives (Notably, kurtosis was the only measure that could not distinguish any cross-language fricative pairs, and the only measure that could not differentiate Cantonese fricatives; it appeared to be the least efficient spectral measure across languages, thus it is excluded from the following discussion). A link was discovered between the size of fricative inventory and the

efficiency of acoustic cues. It appeared that the smaller the native inventory, the more efficient each spectral cue was at differentiating fricatives of that inventory. For Cantonese fricatives, almost all spectral measurements were sufficient to differentiate them. For Mandarin fricatives, CoG, spectral variance, skewness and F2 onset could distinguish all the categories, while peak location failed to distinguish between /f/ and /ʂ/. For English fricatives, spectral variance was the only measurement that could differentiate all the places of articulation.

More importantly, the within-language comparisons revealed that English /f/-/θ/ and /s/-/θ/ contrasts were the least distinguished based on the number of measures that could differentiate them. In addition, for each contrast, different measures failed to differentiate them. It was peak location and CoG that failed at telling /f/-/θ/ apart, while skewness and F2 onset could not separate /s/-/θ/. This finding has shed some light on the differential assimilation of /θ/ between Cantonese and Mandarin. If Cantonese listeners turn out to rely only on spectral cues, while Mandarin listeners depend on both spectral and transitional cues, it could potentially lead to different assimilation of /θ/.

The cross-language comparisons revealed similarities and differences between /f/ and /s/ categories of the three languages. Overall, Cantonese and Mandarin /f/s were different from each other in only 1 measurement: spectral variance, and the /s/s were different from each other in 2 measurements: CoG and spectral skewness. Compared to English /f/, both Cantonese and Mandarin /f/ differed in the same 3 measurements. Compared to English /s/, Cantonese /s/ differed in 3 spectral measurements while Mandarin /s/ differed in almost all spectral measurements. When comparing both Cantonese and Mandarin /f/s and /s/s to English /θ/, there was no acoustic similarities between both /s/s and /θ/, while the /f/s were similar to /θ/ in at least 3 measurements. In other words, the acoustic features of Cantonese fricatives could predict Cantonese listeners' assimilation pattern of English /θ/; however, the acoustic features of Mandarin fricatives could not predict Mandarin listeners' assimilation result of /θ/. These findings indicate that cross-language acoustic similarities may not explain the assimilation of /θ/ by Mandarin speakers.

This study also provided a machine-learning inspired approach that established fricative identifying models with a specific fricative input, which offered some insight to how the fricatives of Cantonese, Mandarin and English were similar or dissimilar in terms of only acoustic information. When applying the models to detect 'native' fricatives, all Cantonese and Mandarin models demonstrated near-perfect accuracy. It is reasonable to argue that Cantonese and Mandarin native fricatives were acoustically distinct. The English /θ/ model demonstrated little accuracy, and this indicates that English /θ/ was particularly difficult to model with the current algorithm compared to the other target fricatives, resulting in a weaker model. The tentative conclusion was that the acoustic information was sufficient for accurate modelling with the current linear algorithm for most of the target fricatives, except for English /θ/.

When applying the models to 'non-native' recordings, the models demonstrated assimilation patterns of the 'non-native' fricatives. Most models performed adequately well, except the models of English /s/ and /θ/. The English /θ/ model could not identify itself with any non-native input for that it was a weaker model. The English /s/ model could only identify less than half of the Mandarin /s/ input which was detected more often by the English /f/ model, indicating that the English models 'considered' the Mandarin /s/ more acoustically comparable to English /f/ than to English /s/. Interestingly, the Mandarin /f/ and /s/ models identified most of the English /f/ and /s/ tokens respectively, in which case it is more in line with the acoustic measurements. There appears to be an asymmetry in terms of the processing of acoustic features across these two languages, which is not explainable by acoustic similarities. This could potentially be a due to the simplified modelling method, and it could lead to interesting further investigation, but it is outside the scope of the current project.

Overall, /f/ and /s/ of each language showed variations in this study, while /s/ showed more variations than /f/. The three /s/s were different from each other in CoG and skewness, and English /s/ and Mandarin /s/ also varied in peak location and variance, meaning that they were different in most of the measurements. This result was supported by the fricative identifier models, as the English /s/ model failed to identify the Mandarin /s/s more than half of the time. One may conclude that Mandarin /s/ and English /s/ were acoustically dissimilar. Meanwhile, Cantonese /s/ varied in 2 measurements from Mandarin /s/, and varied in 3 measurements from English /s/. It

appeared that Mandarin /s/ and Cantonese /s/ were more similar to each other, when they were both different from English /s/. In comparison, the /f/s were less dissimilar to each other, apart from Mandarin /f/ that was significantly different from English /f/ in 3 measurements. The comparison among these three languages revealed that Cantonese and Mandarin /f s/ were extremely different acoustically from English /f s/, while Cantonese /f s/ and Mandarin /f s/ appeared to be slightly different.

As the main research aim of the thesis is to answer the question 'why Cantonese listeners assimilate English /θ/ as /f/ while Mandarin listeners /s/' (see Chapter 1), the current study also compares Cantonese and Mandarin /f s/ to English /θ/. The acoustic measurements show that both Cantonese and Mandarin /f/s were more similar to /θ/ than /s/s, and the fricative models come to the same conclusion. This result is in line with the assimilation pattern of Cantonese listeners. However, surprisingly, neither the acoustic measures nor the Mandarin fricative models could predict Mandarin speakers' assimilation of /θ/ correctly. Previous studies (e.g. Hung, 2000; Zheng & Iverson, 2016) have established the view that Mandarin speakers assimilate English /θ/ to their /s/ category, while the findings of this study mostly point to the direction that Mandarin speakers were more likely to assimilate /θ/ to their /f/ category instead. There was a misalignment between the Mandarin fricative acoustic features and Mandarin speakers' assimilation pattern of /θ/. This misalignment may be an indication for that the assimilation of L2 fricatives is not primarily based on acoustic similarities for listeners with certain language backgrounds.

This finding can be explained by a claim of PAM-L2 (Best & Tyler, 2007), as it hypothesised that the assimilation of L2 sounds is not entirely based on acoustic similarities for L2 listeners. Based on PAM (which is a model established to explain naïve listeners' perception, Best, 1995) and PAM-L2, depending on the experience with L2, naïve Mandarin-speaking listeners and L2 Mandarin-speaking listeners may have different assimilation patterns, and assimilating /θ/ to /s/ may be a specialty of experienced L2 Mandarin listeners. One may argue that listeners' dependence on acoustics similarities may vary according to their different extents of L2 exposure. Notably, in the present study, the fricative identifier models were more closely representative of a naïve listener compared to an experienced L2 listener, as the models were only experienced with one 'native' fricative. Therefore, the models were

identifying and assimilating fricatives based entirely on acoustics. However, real-life listeners with knowledge of more complex phonological systems appeared to make assimilation decisions based on phonological factors other than acoustic similarities, whether they are experienced with the L2 or not. The study by Zheng and Iverson (2016) only tested Mandarin listeners who were naïve to English, and their assimilation of /θ/ still showed a much stronger preference towards /s/ than /f/, despite the acoustic similarities between Mandarin /f/ and /θ/. In addition, the study by Liang (2014) tested university students who had received regular English training throughout school years, and discovered that they still mixed /θ/ and /s/. Therefore, one may conclude that this assimilation pattern is consistent despite the amount of English learning experience, and the acoustic similarities between Mandarin /f/ and English /θ/. This finding supports a shared hypothesis in the SLM framework (Flege, 1995) and PAM-L2 (Best & Tyler, 2007), as it states that perceived similarities on a phonetic level were the primary factor driving perceptual confusion between L1 and L2 sounds. Moreover, the fact that Mandarin listeners consistently perceive English /θ/ as /s/ indicates that the perceived similarity between these two fricative categories is not based on the amount of English training. This leads to a hypothesis that Mandarin listeners that were either naïve to English or had some English training rely on other phonological factors to assimilate English fricative /θ/, while Cantonese listeners rely primarily on acoustic similarities for assimilation.

The focus of the present study was mainly spectral properties of frication, while the only measurement for transition information—F2 frequency at vowel onset—revealed some cross-language comparisons in terms of the potential weighing of formant transitions for fricative processing. The results showed that English /s/ and /θ/ had comparable F2 onset in speech, which may be a potential motivation that leads to Mandarin speakers' assimilation pattern for /θ/ and the difference between Mandarin and Cantonese fricative perception. It is possible that Mandarin speakers would rely more on the F2 onset cue compared to Cantonese speakers. This assumption is in agreement with the conclusion of the study by Wagner et al. (2006) which discovered that listeners whose native language had smaller fricative inventories with no acoustically similar fricatives tend not to use formant transition cues for fricative perception. This agreement could not yet lead to a firm conclusion on the perception

of formant transition. Therefore, further investigation on the perception of formant transitional cues resulting from coarticulation was required from more experiments.

To sum up, spectral properties of the fricatives revealed some cross-language similarities and differences, but the acoustic measurements could not account for the different assimilation patterns of /θ/ in Mandarin and Cantonese. The attempt to model the fricative perception based on acoustic information approximated a simplified naïve listeners' listening situation, and revealed that the complexity that is L2 listeners' fricative processing involves more than just acoustics information. The present study revealed that Cantonese listeners' assimilation of English /θ/ was mainly driven by acoustic similarities, but it did not provide a clear explanation for Mandarin listeners' assimilation decision towards English /θ/. Nonetheless, the findings had shed some light on potential cross-language comparison in terms of cue weighting, which has shown the direction for the next step of research.

Chapter 3 Cross-language Differences in the Perception of Fricative Transitions: Behavioural and EEG Measures

3.1 Introduction

As discussed in Chapter 1, Mandarin listeners tend to choose their /s/ category for English /θ/, i.e., they assimilate English /θ/ to Mandarin /s/, while Cantonese listeners prefer their /f/ category. The previous chapter presented an acoustic analysis of relevant fricatives of the target languages, Cantonese, Mandarin, and English, aiming to investigate possible motivations within the acoustics for assimilating English /θ/ to a specific native category. The results of the study provided some insights into the cross-language acoustic differences among the fricatives with the same phoneme label, but have yet to answer the main research question.

As the study described in Chapter 2 studied a large variety of fricatives produced in natural speech in many different vowel contexts, this meant that it was too complex to analyse transitional cues in detail. The objective from this point became to evaluate the role of transitions from fricatives to vowels in a more controlled context.

As highlighted in the review of Jongman et al. (2000) included in Chapter 2, one measurement of fricative to vowel transition—locus equation—is able to distinguish English /f/ and /θ/, when other spectral measurements cannot. Due to the limitation of the design of the previous study, no measurements were taken regarding the transition from fricative to vowel. It is reasonable to hypothesise that the different assimilation patterns of /θ/ could be relevant to the different roles formant transitional cues play in Cantonese and Mandarin native fricative perception; how listeners perceive and categorise a phoneme could have an effect on how transitional cues are interpreted (Fowler & Brown, 2000; McMurray & Jongman, 2011). Conversely, examining how transitional cues are processed may shed light on why one group of listeners categorise one phoneme differently compared to another group of listeners. Based on the findings, Chapter 2 hypothesised that the different assimilation pattern of /θ/ was relevant to the different roles formant transition plays in Cantonese and Mandarin native fricative perception.

3.1.1 The role of transitional cues in fricative perception

Formant transition is a result of coarticulation, which refers to the fact that the realization of a phoneme segment often varies according to its surrounding phonemes, and the features of this segment often have an impact on a scale beyond its boundary (Kühnert & Nolan, 2009). With coarticulation, the acoustic information of a phoneme is available not only within itself, but also in its transition to the surrounding phonemes. In terms of fricatives, spectral cues within the fricative itself and formant transitional cues are information that are available for fricative perception.

Compared to the spectral cues within the frication part which have consistently demonstrated to have a primary role in fricative identification, dynamic formant transitions are usually considered to be secondary acoustic cues. Indeed, there has been discussion about whether they are essential at all (e.g. Harris, 1958; Jongman, 1989; Jongman, Wayland, & Wong, 2000; Stevens, 2000; Wagner, Ernestus, & Cutler, 2006). Jongman et al. (2000) analysed all the acoustic features of the English fricatives, and discovered that the normalized overall amplitude of the frication could not only distinguish sibilants from non-sibilants, but could also distinguish places of articulation within these two groups. Similarly, analysis of the relative amplitude (the difference between the overall frication amplitude and the overall vowel amplitude) of English fricatives also differentiated all the places of articulation. In other words, providing transitional information did not improve fricative differentiation, as the information within frication sections appeared to be sufficient. This result was in line with the results of an earlier behavioural study by Jongman (1989), in which it was found that hearing an entire fricative-initial syllable did not improve American English speakers' accuracy of fricative identification. Similarly, LaRiviere, Winitz, and Herriman (1975) argued that formant transitions may not contribute to fricative identification—not even for /θ/ which appeared to be the most difficult to perceive among the English fricatives, but they pointed out that transitions may contribute to the perceptual normalisation of speaker variation. On the other hand, findings of other studies have indicated that English speakers do in fact make use of transitional cues when perceiving some fricatives but not others. Harris (1958) revealed that the identification of /f/ and /θ/ was dependent on vowel context; meanwhile, the same group of participants did not make use of transitional information for the identification

of /s/ and /ʃ/. Heinz and Stevens (1961) reached a similar conclusion, as participants' identification of /f/ and /θ/ improved in accuracy when a vowel context was provided, while /s/ and /ʃ/ could be reliably identified from the frication alone. Generally speaking, these studies view perceptual saliency as a main factor in determining the use of transitional cues: only when a fricative is not spectrally distinct, would listeners require the acoustic information contained in formant transitions.

To resolve the conflict between different views on the role of formant transitions in fricative perception, Wagner et al. (2006) conducted a series of experiments with native speakers of not only English, but also Spanish, Dutch, German, and Polish. Among the five target languages, Spanish, English, and Polish have spectrally similar fricatives in their native inventories respectively, while Dutch and German have a relatively sparse fricative inventory with no spectrally similar fricative pairs. The participants heard and identified Dutch or Spanish fricatives. The results showed that compared to Dutch and German listeners, listeners of languages that have spectrally similar fricatives were all more sensitive to misleading formant transitions. The study made 3 major claims: 1) instead of the perceptual saliency, spectral similarity determines the role of formant transitions; 2) listeners' attention to transitional cues is restricted to spectrally similar fricatives; and 3) the listeners apply their native fricative processing strategy when listening to a foreign/unfamiliar realization of a fricative category. With evidence from more languages, this study provided an extensive view of the role of formant transitions in fricative perception.

To date, only one study has tested native Mandarin speakers and their perception of /ʂ/ and /ɕ/, and the result showed that listeners were dependent on both spectral and transitional cues when categorizing the fricative contrast (Mcguire, 2007b). This result provides support for the 3rd claim, and partially for the 2nd claim of Wagner et al.'s (2006) study. The 2nd claim requires evidence from Mandarin speakers' use of transitional cues when perceiving /f/ and /s/. Based on the 3 claims made by Wagner et al. (2006), one might assume that Mandarin speakers would also make use of transitional cues when identifying spectrally similar fricatives (e.g. /s/, /ʂ/, and /ɕ/), but would only rely on spectral information when identifying /f/, the only non-sibilant in their fricative inventory (Duanmu, 2007). One thing worth noting is that Mandarin /f/ has been shown to have a large amount of inter-speaker variability, possibly due to the

lack of a defined spectral peak (Lee et al., 2014). This potentially has an impact on Mandarin speakers' perception of formant transitions. Meanwhile, it is not known how native Cantonese speakers make use of transitional cues. Based on Wagner et al.'s 3 claims (2006), it is more likely that formant transitions should not be necessary for Cantonese fricative perception since there is only /f s/ in the inventory and they are not spectrally similar, as it was shown in Study 1 (Chapter 2).

3.1.2 Phoneme monitoring task

A large number of studies have tested and improved the phoneme monitoring paradigm to facilitate research on perceptual units of speech processing (Connine & Titone, 1996). The paradigm involves a participant listening to a series of speech sounds (e.g. sentences or lists of unrelated speech segments, words or non-words) with target sounds embedded, and pressing a button as soon as a target sound is identified. The task enables detection of phonological similarity or saliency of targets, and provides a measure of participants' sensitivity to certain phonological features (Cutler et al., 1987; Foss & Dowell, 1971; Healy & Cutting, 1976). This paradigm has also been successful in detecting effects of language experience, as task performance in this paradigm appears to be highly correlated with the language background of the listeners (Cutler & Otake, 1994; Finney et al., 1996; Pallier et al., 1993; Wagner et al., 2006). Among these studies, the study by Wagner et al. (2006) shared a similar research interest with the present study. They studied the perception of formant transitions of fricatives using a phoneme monitoring task. The task required the participants to detect fricatives /f/ and /s/ in a search list of trisyllabic pseudowords, with the target fricative always being the syllable initial of the middle syllable. With this task design, they successfully discovered a cross-language difference, which was that fricative cue processing depends on the fricatives' spectral similarity.

Studies using this paradigm showed variations in their choices of dependent variables for analysis, depending on their research interests. The common variables are detection accuracy, reaction time, and rates of false alarms (Connine & Titone, 1996). Foss and Dowell (1971) mainly analysed reaction time, revealing that when target stimuli share more phonological features, the reaction time generally increases in a non-linear manner. The reaction time appeared to be especially longer when the targets sharing

the same phonological features were fricatives. Similarly, McNeill and Lindig (1973) also only measured reaction time, which revealed that the reaction time was minimised when the linguistic level of the target and the search list was the same. On the other hand, the study by Healy and Cutting (1976) and the study by Cutler et al. (1987) analysed both error percentage, which is essentially the same as detection accuracy, and reaction time. Both of the analyses were in line with each other in the two studies, showing the consistency of the effect of the independent variables. Wagner et al. (2006) measured reaction time, and also the frequency of 'timeouts', which means that they analysed how frequently participants were unable to detect a target accurately and/or in time. The latter variable can be considered as a different approach to analysing response accuracy, as it not only includes the number of inaccurate detection responses, but also the time taken to respond such that late but correct responses would not be considered as an accurate response. Notably, whether the result of this analysis was in line with, or was more reliable than the more traditional accuracy analysis was not discussed in the paper. Overall, the fact that studies show flexibility in the choices of the dependent variables in phoneme monitoring tasks, depending on what can best answer their research questions, is an advantage of this paradigm; having more than one dependent variable may increase the reliability of the result.

There is no specific requirement on whether to use one or two button presses in a phoneme monitoring task, in which case researchers can decide how many button presses to include in their task designs, depending on what is the most suitable. Many studies (Cutler et al., 1987; Healy & Cutting, 1976; McNeill & Lindig, 1973; Wagner et al., 2006) ask participants to press a button only when they detect a target sound. Foss and Dowell (1971) adopted the paradigm in their experimental design to use two button responses, a target and a non-target button. No evidence was found showing that there is an effect of number of buttons on the behavioural results.

Another advantage of the paradigm is that it is similar to an oddball paradigm which is widely used to acquire event-related potentials (ERP), especially for components such as mismatch negativity (MMN) and P300 (Luck, 2014). As a result, it is relatively simple to integrate the two, and include both behavioural and electroencephalogram (EEG) measures.

3.1.3 P300 and phonological processing

Despite the many successes the phoneme monitoring paradigm has achieved in deepening our understanding of speech perception, it can only reveal the final stage of perception. We could see what cues may affect phoneme identification by breaking down speech into smaller segments, but it would still be unclear how the decision was made, or at which level of acoustic cue processing the effect had taken place. In other words, there is still a gap in our knowledge about what happens between the moment of hearing a stimulus and the moment when a button is pressed. ERP measure is one way in which an earlier stage of fricative cue processing can be revealed to complement results from behavioural tasks.

ERP has a number of different components representing different levels of brain processing, among which P300 is considered to index inspection and revision of the mental representation induced by target stimuli (Donchin, 1981). It is commonly associated with comprehensive stimulus processing, memory retrieval functions, and updating of working memory (Luck, 2014; Polich & Kok, 1995). Polich and Kok (1995) concluded that P300 could be viewed as a manifestation of neural activities, activated by the processing of the information of the infrequent stimulus, which engages attention to update memory representations. It may reflect a match and a mismatch between the heard stimulus and the stimulus context stored in working memory. In speech perception, working memory is considered important for online processing during conversation (Rönnberg et al., 2013). Due to its limited capacity, it is less engaged when the acoustic input can be easily matched to an existing phonological representation stored in the long-term memory; it is employed to a greater extent when listeners are completing more complex cognitive tasks, especially when speech information is manipulated to create unfavourable listening conditions (i.e. speech with low speech-to-noise ratio, and distorted phonological information etc., Rönnberg et al., 2013; Rönnberg, Rudner, Foo & Lunner, 2008). Thus, through reflecting how much working memory is engaged under various conditions of a behavioural task (e.g. phoneme monitoring task), P300 appears to be an informative indicator that can reveal attention level change as a response to various acoustically manipulated stimuli.

Based on this assumption, studies have employed P300 as an indicator of phonological processing of manipulated acoustic information. Fosker and Thierry (2004) discovered an attentional shift caused by a word-initial phoneme change, demonstrated by P300 modulation in normal adults. Newman, Connolly, Service, and Mcivor (2003) adopted P300 as an index for recognition memory function, with which they argued for perceptual phonological approximation between the absence of a whole consonant cluster and deletion of one consonant within a cluster. Toscano, McMurray, and Dennhardt (2010) combined the measurement of P300 with an auditory oddball task, intending to use P300 measurements to confirm behavioural experiments' suggestion that the acoustic differences of different voice onset time (VOT) were preserved until post-perceptual stages of speech processing. Participants of the study needed to identify a pre-specified target phoneme among filler phonemes and press a button to record their decisions (with one button for "target" and one for "non-target"). EEG was also recorded while the participants were performing the task. Their result showed that P300 amplitude was greater across participants if the stimulus was further away from the VOT boundary i.e. the stimulus that was closer to a prototypical target elicited a bigger P300. Moreover, P300 demonstrated a gradient pattern within one phoneme category depending on the relative VOT distance of the stimulus from the VOT boundary. Therefore, the study argued that P300 was not only sensitive to fine-grained differences, but also sensitive to the phonological categories of the participants. In other words, P300 is affected by both acoustic information and phonological categorisation.

P300 has been shown to be a reliable and sensitive measure of perception in studies of phonological processing, in studies involving various groups of listeners. If P300 is used to investigate acoustic cue processing strategy and cross-language differences in fricative perception, it should be able to provide evidence for how language experience affects phonological processes at an earlier perceptual level; specifically, it will enable the examination of what happens before the listeners' behavioural response in a fricative monitoring task, and potentially to address the different assimilation patterns of Mandarin and Cantonese speakers when listening to the English /θ/.

P300 experiments require a specific design to ensure, and maximize, its elicitation. It is commonly studied using an oddball paradigm, in which two categories of stimuli, a

frequent category (i.e. fillers) and an infrequent category (i.e. targets), are presented in a random sequence. The stimuli that belong to the infrequent category have been shown to elicit a P300 (Polich, 2012), especially when there is a correct behavioural response to a relevant target spotting task towards the infrequent stimuli, i.e. target stimuli (Polich & Kok, 1995). When designing a P300 experiment, apart from using an appropriate paradigm, we should also consider some factors that can affect the magnitude of P300: the overall probability of target stimuli, Target-to-Target Interval (TTI), and attention allocation. The overall probability of target stimuli is calculated as the number of the occurrences of target stimuli over the total number of stimuli presented in a study. Early studies have established that the lower the overall probability of an attended stimulus, the larger the amplitude of P300; in addition, P300 amplitude decreases when there are successive repetitions of a target stimulus (Duncan-Johnson & Donchin, 1977, 1982; Kutas et al., 1977; Squires et al., 1976). Other studies have complemented these findings, and provided us some flexibility to adjust the traditional oddball paradigm to suit specific research needs: these studies pointed out what matters for P300 magnitude was the probability of the task defined stimulus category, rather than of a phoneme category in a study (Katayama & Polich, 1996; Vogel et al., 1998). For example, in a P300 study involving multiple phonemes, each phoneme in the filler category may have a lower probability compared to the probability of the target phoneme, but the target phoneme will still elicit the biggest P300, as long as the target phoneme has a lower overall probability. In Vogel et al. (1998), the absolute probability of the target item was not the lowest compared to any individual non-target item, while it still elicited a much larger P300. Katayama and Polich (1996) indicated that as long as the target stimuli probability was lower than 20%, the probability of each non-target did not have an effect. Notably, despite all being said, a low probability of the target should not be achieved by reducing its number of occurrences. A study indicated that the number of target trials appeared to affect the amplitude of P300; it becomes statistically stable after 20 target trials but changes very little with 30 or more target trials (Cohen & Polich, 1997).

There are changing views on the probability of target category in an experimental task (Luck, 2014; Polich, 2012). The probability of target is calculated as the number of occurrences of a target over the total amount of stimuli presented in the task. Traditionally, studies have argued that the sequential probability of a target stimulus

(i.e., the number of occurrences of the target divided by the total number of stimuli) impacts the P300 magnitude. However, more recent studies have shown that the temporal probability has a greater effect, that is, the probability of the target stimulus occurring over a certain time period. Compared to the traditional view, the recent view on target probability brings a time factor to the sequential probability. Thus, as well as controlling the overall probability of target, TTI should be controlled simultaneously. Gonsalvez and Polich (2002) conducted a study on different TTI in seconds, and its effect on visual and auditory P300 measures. They found out that the longer the TTI, the larger the P300; and for auditory stimuli, when the TTI was longer than 8 s, there was no significant increase in the P300 amplitude. Another study by Polich (2012) demonstrated that when the TTI was 10s long, the probability effect disappeared, which means that the P300 amplitude was comparable when target probability was 80% and when it was 20%. Evidence shows that the TTI is a crucial factor in the designing of a P300 experiment when attempting to ensure a larger P300 amplitude. It is preferable to maintain the TTI at over 8 s.

In addition, P300 appears to be attention-driven, which means that the amount of attentional resources allocated for completing the task will modulate the magnitude of P300 (Polich, 2007). Active stimulus processing (e.g. doing an identification task while listening to the stimuli) can produce larger P300 responses, but for tasks that require a very large amount of attentional resources, i.e. the tasks are too challenging, P300 amplitude will be smaller and peak latency longer as more attentional resources are allocated to performing the task (Kok, 2001). In the study by Fosker & Thierry (2004), P300 modulation was used as an index of attentional shift of normal and dyslexic adults, and revealed that the dyslexic group was not paying attention towards certain phonological cues during the experiment. The study also argued that the dyslexic group had a deficit in phonemic cues awareness rather than between-category discrimination. This study is an example of adopting P300 in showing differences in phonemic sensitivity between groups.

To conclude, P300 is a reliable and informative index of phonological processing. It appears to be sensitive to the phonological categories during phonological processing, and it preserves some effects of acoustic details. Using P300 to investigate perceptual strategy and cross-language differences in fricative perception should therefore

provide evidence of how language experience affects phonological processes, and lend support to behavioural results.

3.1.4 The aim and design of the present study

The aim of the present study is to investigate possible cross-language differences in the weighing of formant transition cues during fricative perception by Mandarin, Cantonese and English speakers, in order to understand the assimilation patterns of the English /θ/. This study was designed to combine the benefits of a phoneme monitoring paradigm and P300 measurement, intending to explore the effect of formant transitional cues on both an earlier level and the final level of fricative identification. The study integrated a fricative monitoring task and traditional oddball paradigm into an active 'oddball monitoring' paradigm, in which a series of monosyllabic, and CV-structured stimuli were each identified as "non-target" or a relatively infrequent "target". The non-target stimuli consisted of fricatives of Cantonese, Mandarin and English other than /f/ and /s/. The target stimuli, initially /f/ or /s/, were either cross-spliced (i.e., frication spliced into a vowel context from another fricative, such as /f/ replacing the frication in /sɑ/) or identity spliced (i.e., /f/ replacing the frication of a /fɑ/ syllable), such that the splicing operation was the same for the two types of stimuli. The identification of targets was measured both behaviourally (i.e., accuracy of identification and reaction time of button presses) and in terms of the P300 ERP from EEG recordings.

The present study only used /f/ and /s/ as target fricatives, which are shared fricative categories by Cantonese, Mandarin, and English. In a previous study examining native English speakers' perception of foreign fricatives, /ʂ/ and /ɕ/, listeners relied completely on formant transitions to differentiate the two, and ignored the differences in the fricative spectra (Mcguire, 2007a). This finding contrasts with Wagner et al. (2006), which showed that English listeners relied mainly on fricative spectra to process native fricatives. Such a contrast indicates that when presented with fricative categories that do not exist in the native inventory, listeners may adopt a listening strategy that does not necessarily resemble their native fricative processing strategy. In order to examine native fricative perception, the present study does not include /θ/ as a target, as this does not exist in either Cantonese or Mandarin.

Based on the theories and evidence discussed above, cross-language differences in the perception of formant transitions during fricative processing are likely correlated to the differences in native fricative inventories (e.g. Wagner et al., 2006). If a language has pairs of acoustically and perceptually similar fricatives in its inventory, it is likely that its speakers would attend to formant transitional cues. Therefore, it is hypothesised that Cantonese speakers will be less attentive, or show no attention, to formant transitions compared to Mandarin speakers and English speakers, due to Cantonese's small fricative inventory and its members' distinct acoustic features. It was expected that Cantonese speakers would perform similarly under both cross-spliced and identity-spliced conditions in the phoneme monitoring task, and also demonstrate a comparable P300 under both conditions. Meanwhile, as Mandarin and English have more complex inventories with acoustically similar fricatives, it is expected that the behavioural task performance of Mandarin speakers and English speakers would deteriorate under the cross-spliced condition, and their P300 in this condition would also be smaller in magnitude, as their fricative processing was expected to be affected by misleading formant transitions.

3.2 Method

3.2.1 Subjects

This experiment tested 12 native Southern British English speakers (7 females and 5 males), 12 native Northern Mandarin Chinese speakers (9 females and 3 males), and 12 native Hong Kong Cantonese speakers (11 females and 1 male), who were all right-handed adults between 18 to 30 years old. A pre-test questionnaire (see Appendix B) was given to each participant who reported no history of hearing, learning, or language impairment, and no history of neurological disorders.

The native English speakers were all monolinguals from birth, and they had no knowledge of either Mandarin or Cantonese. The Mandarin-speaking and Cantonese-speaking participants were also monolinguals from birth; however, they had English learning experience and were exposed to an English-speaking environment as they were students studying in London. To limit the effect of English exposure, the participants satisfied specific criteria: the Mandarin speakers had started taking English classes at school no earlier than 6 years old, had been exposed to an English-

speaking environment for less than 2 years, and had no experience of Cantonese; the Cantonese speakers had started learning English no earlier than the age of 5, had been living in an English-speaking environment for less than 2 years, and did not claim to be fluent in Mandarin Chinese in the questionnaire (see Appendix B).

All participants passed a hearing screening using a 2-channel pure tone audiometer; they were able to detect a tone at any frequency from 250 to 8000 Hz 100% of the time at 25 dB SPL with either ear.

3.2.2 Stimuli

Four female native speakers of each target language produced the stimuli in their native language for this experiment. The English speakers were instructed to pronounce the stimuli as English syllables, and the Cantonese and Mandarin speakers read Chinese characters. Their production was recorded at 44100 Hz 16-bit sample rate, and the processing of the recordings was conducted in Praat (Boersma & Weenink, 2017). The speakers read out native fricative-vowel syllables listed in Table 3-1 (V were only low vowels /a ɑ/, as shown in Table 3-1), each for 20 times, and 5 of each syllable were selected to be the stimuli based on the quality of production. A 100 Hz high-pass Hann band-pass filter was applied to the recordings, with smoothing frequency at 50 Hz. The amplitude was normalised to 70 dB SPL across stimuli. The length of each stimulus was equated to 550 ms, with around 150 ms of frication and 400 ms of vowel. For further manipulation, the fricative and vowel boundary of each selected syllable was marked at the zero-crossing point at the end of frication and the beginning of harmonic structure, as shown in Figure 3-1. The target syllables were spliced in two ways, identity spliced and cross spliced. An identity-spliced stimulus had its fricative replaced by the same fricative of another token from the same language. A cross-spliced fricative was replaced by the other target fricative of a token from the same language (e.g. the /f/ of an English /fɑ/ was replaced by a /s/ from an English /sɑ/). The point of splicing was the marked boundary stated above, following the method adopted by Wagner et al. (2006). The filler syllables were only identity spliced. All the splicing took place within talker.

The stimuli were presented in 6 blocks, 3 with /s/ and 3 with /f/ as the target. As shown in Table 3-1, each block had 60 target trials, yielding 180 target trials in total per target

per participant. To maximize the magnitude of P300, the TTI was 9 s and above (vary depending on the response time), achieved by adding 2 to 4 fillers randomly in between targets. In each block, the probability of target stimuli was under 25%, and depending on the participants' perception, the perceived probability was expected to vary between 12.5% and 25%. The stimuli sequence of each block was randomly generated each time.

Table 3-1 Syllables used as stimuli in the active P300 experiment.

	Cantonese		**English**		**Mandarin**	
Target syllables	f s	a	f s	ɑ	f s	a
Filler syllables			θ ʃ v z ð ʒ	ɑ	ʂ	a

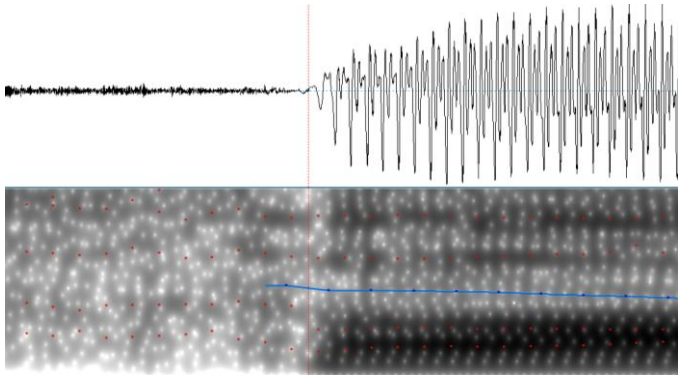


Figure 3-1 The point of splicing shown in Praat.

3.2.3 Apparatus

The stimuli were presented using Praat (Boersma & Weenink, 2017), through an RME Fireface UC audio interface, and Etymotic ER-1 insert earphones. The time-aligned triggers for the stimuli were generated and recorded as pulses on a disused audio channel, which were then converted to transitor-transitor logic (TTL) triggers using a custom circuit to record the stimulus onset time. Behavioural responses, i.e. button presses, were recorded using a custom button box with two buttons (one for targets and one for non-targets) to register accurate reaction time. The triggers, button responses, and EEG signals were recorded using a Biosemi Active Two system with

64 pin-type active-electrodes plugged into electrode holders on a Biosemi headcap, and 7 external flat-type active-electrodes (left and right mastoids, nose, two vertical and two horizontal EOG electrodes for eye movements). The sampling frequency of the EEG recordings were 2048 Hz. Impedance of each electrode was controlled within the range of ±25kΩ during the testing sessions of all participants. All the testing was conducted in a sound-proof, electromagnetic shielding booth, with the participants sitting in an armchair holding the button box in their hands.

3.2.4 Procedure

The participants were informed what the target phoneme would be, and that it would be embedded in language-nonspecific CV syllables before the start of each block. The three blocks with the same target, with short breaks in between, were always presented in a row, followed by a longer break to help the participants prepare for the change of the target. The order of target block groups was counterbalanced among participants. The target phonemes were printed on separate sided of an A4 paper stuck on a wall in front of the participants, reminding them what target to listen for during testing. Two button press responses were required. They were asked to press the target button (marked with a sticker) on the button box as soon as they identified a target stimulus, and to press the other button for any other stimuli. This was to eliminate the potential motor difference that may appear in the EEG signal, and consequently contaminate the data. They were informed in advance that the next sound would play shortly after they had pressed a button. Their performance was monitored outside the testing booth.

3.2.5 Analysis

3.2.5.1 Pre-processing of the EEG data

Pre-processing of the EEG data was conducted offline using MATLAB with plug-in toolboxes. The EEG recordings were referenced to the average of the two mastoids, and they were high-pass filtered at 0.1 Hz and low-pass filtered at 40 Hz using Butterworth filters, implemented by the ERPLAB plugin (Lopez-Calderon & Luck, 2014) within EEGLAB toolbox (Delorme & Makeig, 2004). With the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011), noisy channels were interpolated, and Independent Components Analysis was conducted for correcting ocular movements.

3.2.5.2 P300 analysis

After pre-processing, the recordings were epoched with 200 ms before the stimulus onset and 1000 ms after the stimulus onset intervals. Trials with an incorrect behavioural response were excluded. The selected epoched trials were cleaned using Denoising Source Separation (DSS; de Cheveigné & Simon, 2008), which increased the signal-to-noise ratio of the neural data.

Following the study by Toscano et al. (2010), P300 magnitude was calculated in the present study as the mean voltage between 300 ms and 800 ms after the stimuli onset, averaging across the parietal and mid-line channels, 15 in total.

3.2.5.3 Behavioural data analysis

Behavioural performance of each participant was analysed by the percentage of targets identified, indicating detection accuracy, and the median reaction time, calculated as the time difference between a button press and a stimulus onset under each condition. Then the mean of both measures was calculated across participants.

3.3 Results

3.3.1 Behavioural results

Table 3-2 shows the descriptive statistics of the behavioural results.

3.3.1.1 Detection accuracy

The detection accuracy was indicated by the percentage of correct target identification, and it revealed that all language groups were affected by cross splicing, with target identification accuracy dropping to about chance level for the cross-spliced stimuli, as shown in Table 3-1. A mixed-design analysis of variance (ANOVA) was conducted; there was a significant main effect of *stimulus type*, $F(1,33) = 215.33$, $p < .001$, with more target identifications for identity-spliced than cross-spliced stimuli. There was a main effect of *participant language*, $F(2,33) = 4.44$, $p < .05$, with English speakers being slightly more accurate when identifying targets. There was a significant interaction between *target* and *stimulus type*, $F(1,33) = 22.60$, $p < .001$; identification of /f/ was more affected by cross splicing than /s/. However, there was no significant

interaction involving *participant language*, $p > .05$, indicating that the effect of transitions was similar between the language groups.

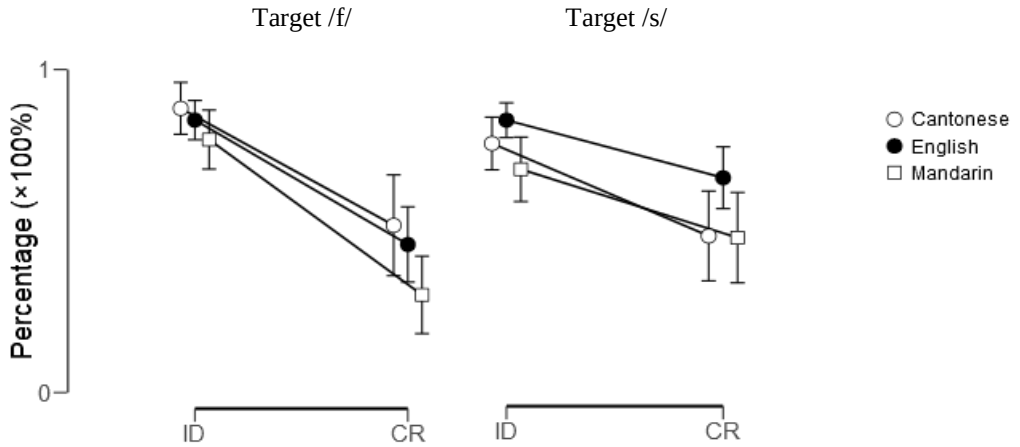


Figure 3-2 Average percentages of correct target identification for participants of three language groups under different conditions. ID = identity-spliced stimulus type, CR = cross-spliced stimulus type.

3.3.1.2 Reaction time

The result of a mixed-design ANOVA showed that the only significant main effect was *stimulus type*, $F(1,33) = 15.52$, $p < .001$. That is, listeners were slower at identifying the target for cross-spliced stimuli, even when considering only trials in which a target was identified. All the interactions involving *participant language* were non-significant, $p > .05$.

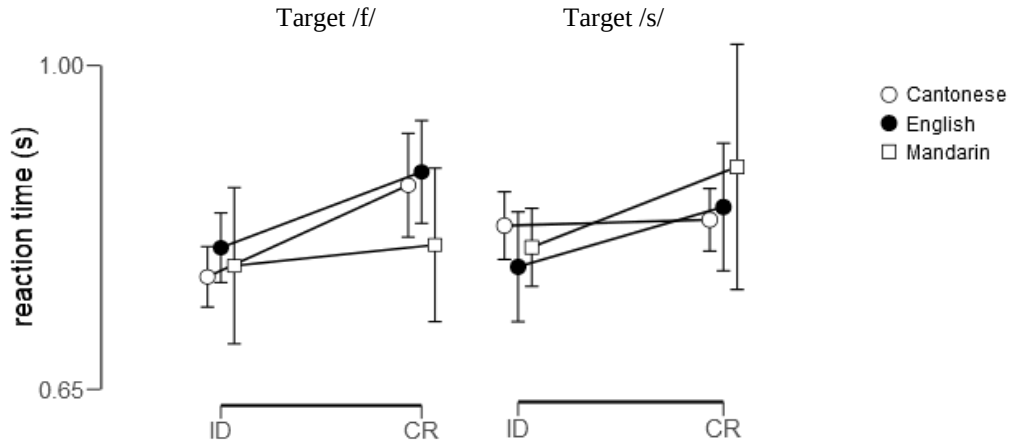


Figure 3-3 Average reaction time of participants of three language groups under different conditions. ID = identity-spliced stimulus type, CR = cross-spliced stimulus type.

Table 3-2 Descriptive statistics of behavioural results for each language group across subjects. ID = identity-spliced, CR = cross-spliced.

Target fricative	Stimulus type	Subject language	Mean detection accuracy (×100%)	Standard deviation	Mean reaction time (s)	Standard deviation
/f/	ID	Cantonese	0.880	0.079	0.772	0.145
		English	0.843	0.082	0.803	0.134

Target fricative	Stimulus type	Subject language	Mean detection accuracy (×100%)	Standard deviation	Mean reaction time (s)	Standard deviation
		Mandarin	0.783	0.180	0.784	0.091
	CR	Cantonese	0.518	0.259	0.871	0.199
		English	0.458	0.216	0.885	0.198
		Mandarin	0.302	0.196	0.806	0.189
/s/	ID	Cantonese	0.766	0.151	0.811	0.148
		English	0.838	0.093	0.770	0.164
		Mandarin	0.685	0.238	0.789	0.182
	CR	Cantonese	0.479	0.248	0.817	0.166
		English	0.659	0.186	0.830	0.193
		Mandarin	0.473	0.206	0.871	0.296

3.3.2 P300 results

As shown in Figure 3-5, P300 was elicited when there was a target button press towards a target, and was not elicited when there was a non-target button press towards a filler. A mixed-design ANOVA was conducted including the filler *stimulus type*, with *participant language* as a between-group factor, and *stimulus type* and *target* as within-group factors. There was a significant main effect of *stimulus type*, $F(2,66) = 43.81$, $p < .001$. Post-hoc tests using the Bonferroni correction revealed that the mean amplitude of P300 in the filler condition was significantly lower than that in the identity-spliced and the cross-spliced target stimulus conditions, with a mean difference of 2.31 and 1.85 respectively, $p < .001$. The detailed descriptive statistics are shown in Table 3-4. This suggests that in the target conditions, P300 responses were significantly bigger than those in the control condition (responses to filler stimuli).

To investigate the difference between the target conditions, a mixed-design ANOVA was conducted excluding the filler responses from *stimulus type*. The only significant main effect was *stimulus type*, $F(1,33) = 7.69$, $p < .01$, indicating greater mean P300 amplitude for identity-spliced stimuli across languages. All the interactions involving *participant language* were non-significant, $p > .05$, which means that the P300 measures showed no cross-language differences.

Although not statistically significant in the analysis, it is noticeable that the P300 responses to /s/ of the Mandarin-speaking participants were comparable across

identity-spliced (red line, Figure 3-5) and cross-spliced (blue line, Figure 3-5) conditions (see

Table *3-3* and Figure 3-5).

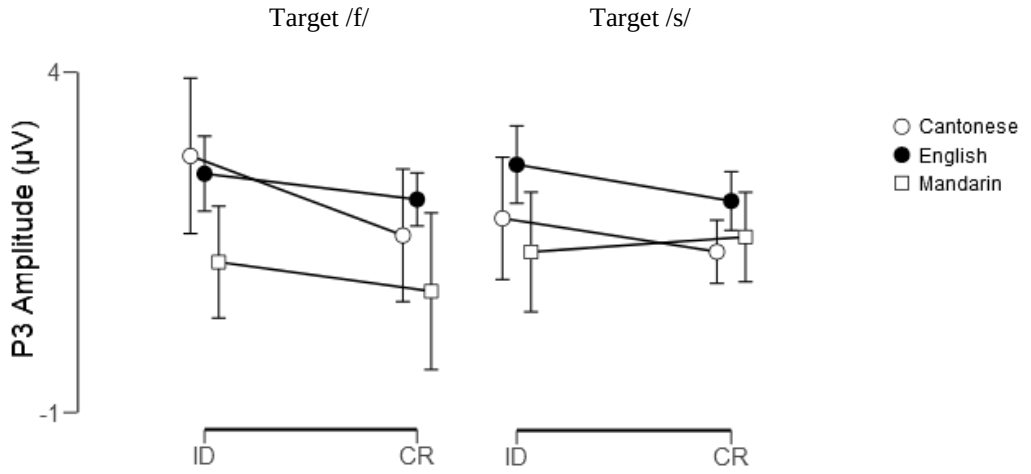


Figure 3-4 Average P300 amplitudes of three groups of participants under different conditions. ID = identity-spliced stimulus type, CR = cross-spliced stimulus type.

Table 3-3 Descriptive statistics of P300 measurement for each language group across subjects. ID = identity-spliced, CR = cross-spliced.

Target	Stimulus Type	Subject language	Mean P300 amplitude	Standard deviation
f	ID	Cantonese	2.769	2.677
		English	2.508	1.621
		Mandarin	1.210	2.097
	CR	Cantonese	1.602	1.113
		English	2.130	1.253
		Mandarin	0.781	3.189
	filler	Cantonese	-0.589	1.607
		English	0.521	1.253
		Mandarin	-0.883	1.451
s	ID	Cantonese	1.858	2.433
		English	2.645	1.142
		Mandarin	1.365	1.633
	CR	Cantonese	1.369	2.001
		English	2.112	1.018
		Mandarin	1.585	1.321
	filler	Cantonese	-0.413	1.986
		English	0.639	1.185
		Mandarin	-0.797	1.269

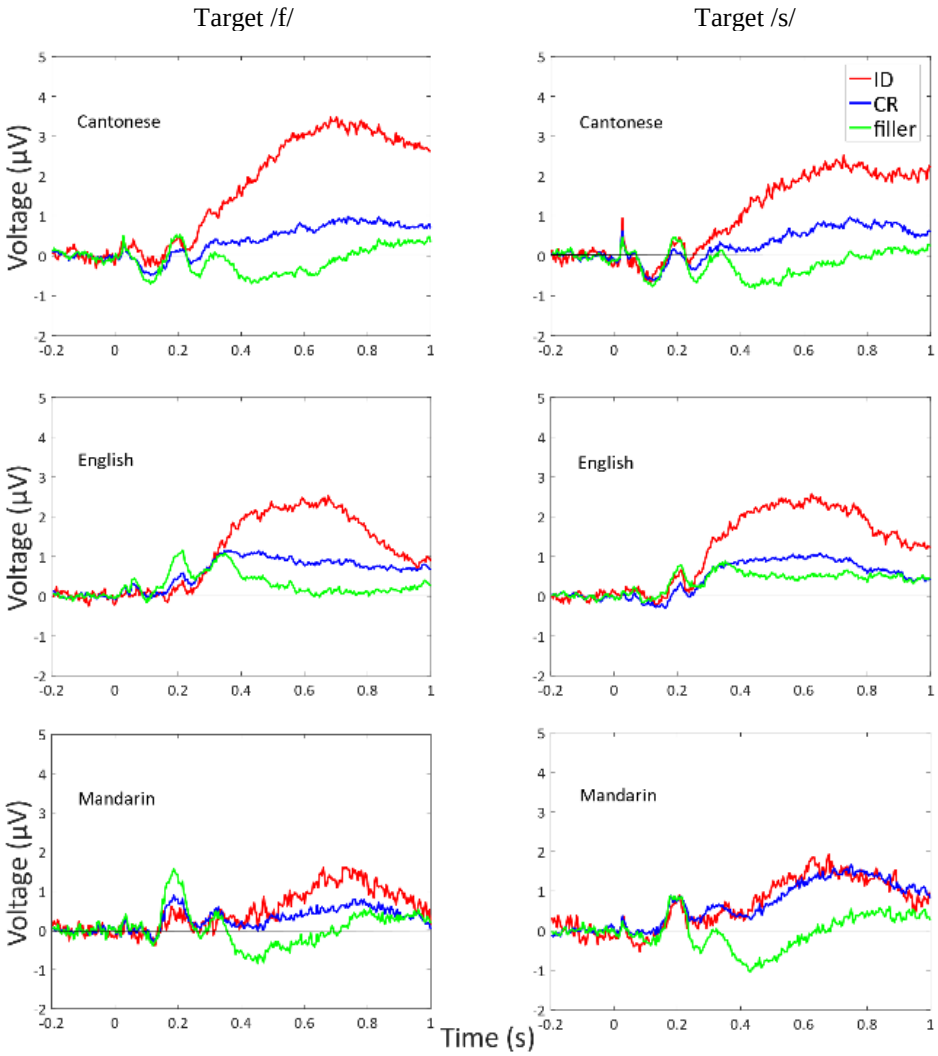


Figure 3-5 Grand-average ERP waveforms, averaged across parietal and mid-line electrodes as a function of stimulus type and subject language. ID = identity-spliced stimulus type, CR = cross-spliced stimulus type.

3.4 Discussion

The current study investigated the potential differences among native Cantonese, Mandarin, and English listeners in terms of their sensitivity to transitional cues during fricative processing. Overall, all participants' performance deteriorated under the cross-spliced stimulus condition. They were all sensitive to the mismatching formant transitions induced by cross-splicing target fricatives and vowels despite their different language backgrounds. The effect was demonstrated by a significant main effect of stimulus type in both behavioural and EEG measures, and no significant interactions involving participants' language; more specifically, lower accuracy, longer reaction time, and a smaller mean P300 amplitude under the cross-spliced condition.

The P300 results demonstrate that the manipulated acoustic information, i.e. the mismatched formant transitions, was preserved at a later stage of perception of the participants, which is comparable to previous P300 results (Fosker & Thierry, 2004; Toscano et al., 2010). P300 amplitude is considered to vary with the amount of attentional resource allocated in the task (Polich, 2007, 2012); the lower amplitude under cross-spliced conditions across language groups may indicate that their attention was divided after a cross-spliced stimulus, as the listeners may have continued to search for additional information elsewhere to resolve conflicting cues from the stimuli. Alternatively, a smaller P300 amplitude towards a type of stimuli can also indicate that listeners judge them to be poorer realizations of a phoneme category (Toscano et al., 2010); in other words, the stimuli with mismatching formants were spotted, and not deemed good enough exemplars for the listeners to update the fricative representation in their working memory.

The finding that English and Mandarin speaking participants were affected by misleading formant transitions is mostly in line with the initial hypotheses. The English participants behaved similarly as the ones in Wagner et al.'s study (2006), as they attended to the mismatched transitions for both /f/ and /s/; moreover, the behavioural accuracy result suggests that the effect of mismatching transitions was greater for /f/ than for /s/. It was also expected that Mandarin speakers would use formant transitions in identifying /s/ due to their experience with spectrally similar fricatives; they have a denser fricative inventory, especially in the area between the alveolar and the palatal regions. The behavioural measures supported this hypothesis. However, it was surprising that the mismatching transitions were not reflected in the Mandarin group's P300 responses for /s/. Though this was not statistically significant, there was evidence for this trend shown by the grand-average waveforms (see Figure 3-5). This potentially revealed the Mandarin speakers' higher level of tolerance towards various realizations of /s/ transition. Prior to the study, it was unclear whether the mismatching transitions would also impede the Mandarin group's identification of /f/, since it is Mandarin's only non-sibilant and, as demonstrated in Chapter 2, is spectrally distinct from other sibilants in the fricative inventory of Mandarin. However, the current results showed that the mismatch effect was greater for /f/ than for /s/. One possible explanation is related to the between-speaker variability of Mandarin /f/: a study has shown that Mandarin /f/ is more spectrally varied than other

fricatives (Lee et al., 2014). It is possible that listeners need to make use of transitional cues since spectral cues are not always consistent.

Contrary to the hypotheses, just like the English and Mandarin participants, the Cantonese-speaking listeners also attended to the mismatching formant transitions, despite having a much smaller fricative inventory and no spectrally similar fricatives. Similar to Dutch, Cantonese has only two fricative categories /f/ and /s/, so theoretically native speakers do not need to attend to coarticulatory cues for fricative perception as the fricatives are spectrally distinct (Chan & Li, 2000; Wagner et al., 2006). The result of the present study challenges the generalisability of the view of Wagner et al. (2006) on the relevance between the use of coarticulatory cues and native fricative inventory, as it would predict that the Cantonese participants would not attend to mismatching formant transitions due to their small fricative inventory and its dissimilar fricatives.

3.4.1 The primary unit for speech processing

The conflict between these results and the conclusions of the previous studies indicates that some factors, other than fricative inventory density and spectral similarity, may affect the use of formant transitional cues. One important factor might be the primary unit for speech perception and production, which cannot be revealed by the present phoneme monitoring experimental design (McNeill & Lindig, 1973). Previous studies on Mandarin have claimed that syllables are stored and retrieved as a whole, and that syllables, instead of phonemic segments, play a primary role in speech processing and production (Chen, 2000; Chen, O'Séaghdha, & Chen, 2016). Storing and retrieving all the syllables as segments is not an impossible mission for Mandarin speakers. Setting tone variations aside, there are only 416 syllables in Mandarin, among which 399 are commonly used (Da, 2010; Chinese Academy of Social Sciences, 2018). As a result, the frequency of use of each Mandarin syllable is relatively high. The fact that each Mandarin syllable is highly frequently used in daily speech makes storing and retrieving each syllable a simpler task, and likely a more efficient way than processing each phoneme segment within those syllables every time they occur. Frequency of use is also considered to link to linguistic form reduction (Bybee, 2001), as the higher the frequency, the more drastic the reduction. This finding might also apply to syllable

structures, as evidence in Mandarin has shown a clear reduction of features in both coda and onset positions of a syllable, while the nucleus is usually preserved, or with additional features (Duanmu, 2007). Another study has shown reduction in syllable duration in Mandarin depending on syllable position, such that the duration of a medial syllable can be less than half of the average duration of a one-syllable phrase (Xu & Wang, 2009). There is a large range of ways in which syllables can be reduced in Mandarin (Burchfield & Bradlow, 2014; Duanmu, 2007), and so formant transitions, a cue that is usually preserved, could become crucial for syllable identification. Consequently, Mandarin speakers would attend to the formant transitional cues despite the phoneme context.

Cantonese speakers may attend to formant transitions for similar reasons. Cantonese, irrespective of its tones, has 750 syllables, including commonly mispronounced versions of standard syllables (these are included in the total number due to their frequent appearance in daily speech, which is caused by lack of standardization of Cantonese, and which is different from Mandarin which is standardized and highly regulated) (Bauer & Benedict, 2011). Although the number of syllables in Cantonese is significantly higher than the number of Mandarin syllables, it is still far smaller than the number of English syllables. A smaller number of syllables means that each will be used with a higher frequency. Based on the argument above, Cantonese speakers attending to formant transitions is an explainable outcome. Previous studies have provided evidence for Cantonese speakers treating a syllable as the unit which is activated primarily during speech perception; in other words, Cantonese speakers perceive and process syllables as a whole (Chen & Yip, 2001; Cheung & Chen, 2004; Wong, Huang, & Chen, 2012; Wong, Wang, Wong, & Chen, 2018). This could lead to their different use of formant transitions in perception. A comparative study of holistic processing of syllables by Cantonese and English speakers (Liu & Hsiao, 2014) observed that Cantonese speakers' perception of the syllable-initial and -final phonemes appeared to be influenced by surrounding phonemes within that syllable, but that these had little influence on English speakers' perception. The Cantonese speakers tended to combine features of a syllable to form an all-inclusive representation, and showed difficulty in selectively attending to individual phoneme segments. This indicates that formant transitions are stored within the syllable representations, to which syllables with incongruent transitions could not match. These

findings support the results of the current study, as they offer an explanation for why Cantonese participants also attended to the mismatching formant transitions, despite having a small fricative inventory.

Another theory that may also support the view on Cantonese and Mandarin speakers discussed above is the potential effect of early orthographic experience and its impact on syllable processing. Cheung and Chen (2004) showed that non-alphabetic learners, i.e. logographic learners like those of Cantonese and Mandarin, showed little phonemic-level analysis during speech processing compared to alphabetic leaners. The perceptual integrity of syllables appeared to be enhanced by the learning of logographs, in this case, the Chinese characters. In their study, exposure to Pinyin was considered as a type of alphabetic learning experience, and the younger generation in mainland China were mostly exposed to Pinyin and Chinese characters during early years of school education. This was different from the Hong Kong Cantonese speakers, who were only learning logographs without Pinyin. Whether this unique early exposure to both alphabetic and non-alphabetic language learning experience has led to a different syllable-processing strategy in Mandarin listeners is unclear, as the current study showed sensitivity to coarticulatory cues that may be related to either syllable-level processing or phonemic properties.

The findings about cross-linguistic differences in the primary perceptual units of speech may account for the behavioural result and most of the P300 results, but it is less clear how they can account for Mandarin listeners' comparable P300 responses in the different stimulus type conditions for /s/. Although this requires further evidence as it this was only a trend and was not a statistically significant effect, one possible interpretation is that this indicates that the Mandarin group's attention to formant transitions served as a cue during syllable recognition and categorisation; once a syllable was categorised to the /sa/ category, the listeners shifted their attention away from the formant transitions, no matter whether they matched or mismatched. The Mandarin group showed higher levels of tolerance towards mismatching formants within the categorised /sa/ syllables. This may be because in comparison to Mandarin /f/, Mandarin /s/ is a more flexible category, tolerant to a wider range of variation, such that a greater number L2 fricatives are assimilated to it.

To summarize, the participants in the current study were informed that the target stimuli were not language-specific, but they were familiar categories in their native languages, and it is likely that they approached the task differently. To be specific, the English listeners were identifying syllable-initial /f/ and /s/, while the Mandarin and Cantonese listeners were in fact looking for /fa/ and /sa/ syllables. The outcome was that they were all sensitive to the misleading formant transitions but for different reasons. Current theories of fricative perception have mainly focused on phonemic-level processing, while the discussion above argues that the perception of cues during fricative processing appears to involve more factors than just fricative inventory density and spectral similarity. Thus, based on the discussion above, future research studying speech perception of Mandarin speakers and Cantonese speakers should take their syllable processing strategy into consideration.

3.4.2 Limitations and conclusions

Chapter 1 established that there are differences in perception of the English /θ/ by Cantonese and Mandarin native speakers at a behavioural level, and the research aim of the project is to discover what leads to the differences. The current study revealed that these differences do not lie in differences in sensitivity to mismatching formant transitional cues, and thus did not provide a conclusive answer to the initial research question, which aimed to address why there is a difference in perception, and at which level of processing this difference occurs. The discussion above argued that the Cantonese and Mandarin speakers treated a syllable as an integral perceptual unit, so they attended to the formant transitions. However, the result does not rule out the possibility that the different perception of /θ/ is linked to their different use of coarticulatory cues. Due to potential differences in syllable perception, we should shift our focus from processing of specific cues to the interactions among cues within a syllable. It is possible that differences in the perception of /θ/ are caused by an interactive effect that includes both spectral properties of the fricatives and the formant transitions. Changing only one of these cues may have meant that it was not possible to reveal the effect of the interaction of the cues; a finer-grained manipulation of the stimuli, which could generate step-by-step changes in transition may help us discover such an interaction.

Admittedly, as all the data collection was conducted in London, the participants' English experience may have altered the participants' strategy of using formant transitions to some extent. Currently, no other studies have provided sufficient evidence to determine how much training is needed for L2 listeners to learn to make more use of formant transitions during L2 learning. Although participants had only lived in an English-speaking environment for a maximum of 2 years in an attempt to limit the effect of English learning experience, it was difficult to eliminate.

A complete mismatch of fricatives and vowels, without spectral smoothing, may sound unnatural to most speakers in a quiet sound-proof booth with high resolution audio delivery platform. The splicing procedure of this study's stimuli replicated the method adopted by the study by Wagner et al. (2006). But compared to their study, which embedded the cross-spliced syllables within trisyllabic pseudo-words, the phoneme monitoring task in this study might have been less perceptually demanding for listeners. In this case, it may have been relatively easy for listeners to have noticed the mismatching formant transitions. Although it was a necessary change to fit in the oddball paradigm for P300 measuring, the current study is not a full replication of Wagner el al.'s study (2006), and so it not possible to directly compare the results of these studies.

In conclusion, the current study suggests that the processing of fricatives may involve more elements than just phoneme inventory density and spectral similarity of the fricatives, depending on the language in question. In the case of Cantonese and Mandarin, their respective speakers appeared to attend to formant transitions, despite the size of their native fricative inventory, and the spectral similarity among their native fricatives. Given evidence from previous studies, this suggests that the perception of fricatives may be affected by how they perceive syllables. To study assimilation of English /θ/ then, a complete mismatch of formant transition may not be sufficient, as finer-grained interaction among the cues within a syllable may influence processing of the different cues. The following chapter will examine the influence of a more detailed manipulation of fricative-initial syllables on perception, with the aim of discovering the cause of the different assimilation pattern for /θ/.

Chapter 4 Cross-language Differences in Detailed Fricative Cue Weighting during Perception

4.1 Introduction

The study presented in the previous chapter investigated how listeners from different language backgrounds differed in their identification and perception of mismatched fricatives and vowels. Specifically, it tested how cross-splicing /fɑ/-/sɑ/ and /fa/-/sa/ affected the accuracy and reaction time of identification, and the magnitude of P300, of the native Cantonese speakers, the native English speakers, and the native Mandarin speakers. In general, the study revealed that all the subjects were sensitive to the incongruent transitional cues to some extent, and no significant cross-language differences were found. It was surprising that the Cantonese speakers were also sensitive to the cross- spliced stimuli, despite their relatively sparse fricative inventory. However, some studies have shown that, different from English speakers, Cantonese and Mandarin speakers may process syllables as a whole unit, and as a result they may have a higher level of sensitivity to transitional cues (e.g. Chen & Yip, 2001; Chen, 2000). Therefore, a complete mismatch of fricative and vowel may catch Cantonese and Mandarin speakers' attention more easily, despite the sparseness of their fricative inventories and the degree of spectral similarity fricatives between them.

The issue raised in Chapter 3 is that the perceptual difference between Cantonese and Mandarin speakers exists when categorising English /θ/, and it cannot be traced back to any difference in sensitivity to a complete mismatching formant transition from the fricative to vowel. The study shifts the focus from fricative categorisation to syllable categorisation, with finer-grain acoustic manipulation and mismatch. This study would like to answer some questions that emerged from the results of the previous chapters: 1) if all listeners are sensitive to a complete transition mismatch, do they perform differently when the mismatch is finer-grained; in other words, do they rely on transitional cues to different extents, and 2) do native syllable boundaries differ for listeners of different native languages, and if so, how do they affect the identification of an L2 syllable?

One way to address these questions, together with the main research question of the thesis, could be to use an approach that integrates fine-grained manipulation of two groups of acoustic cues (i.e. frication cues and transitional cues), and create step-by-step mismatches within syllables. This method has been shown to be insightful, showing details that were difficult reveal by a complete mismatch of a vowel and a fricative. For example, Yu and Lee (2014) created a /s/-/ʃ/ continuum of stimuli with step-by-step change in the frication within a syllable, and cross-spliced these with the vowels /a/ and /u/ which were taken from naturally produced American English /da/ and /du/[1]. They conducted an experiment with these stimuli revealing perceptual compensation for coarticulation within fricative syllables in different vowel contexts, demonstrating the sensitivity of a two-alternative forced choice (2AFC) task with stimuli that had fine-grained stepped change in spectral cues. The result revealed a significant shift of fricative boundary when the vowel context was different: more tokens were identified as /s/ when the vowel was /u/, indicating a perceptual compensation effect for coarticulation.

Another study investigated sensitivity to step-by-step change in both the frication cues and the vowel cues. Mcguire (2007b) carried out a fricative assimilation study in native Mandarin and native English speakers, using Polish voiceless alveolo-palatal and voiceless retroflex fricatives. The study separately interpolated the fricative portion and the vowel portion of naturally produced Polish syllables: /ɕa/ and /ʂa/, and created a 10×10 stimuli grid, where 10 frication steps on the fricative continuum were cross-spliced with 10 vowel steps on the vowel continuum. Participants from each language group performed the same phoneme labelling task by pressing the respective buttons. Since both target fricatives were native to the Mandarin speakers and non-native to the English speakers, cross- language differences in syllable boundary were expected. The results demonstrated that Mandarin-speaking subjects' syllable boundary was placed diagonally across the stimuli grid, which indicated that they relied on both fricatives and vowels to label the two syllables; on the other hand, the English-speaking subjects only relied on the vowel step change, giving a boundary perpendicular to the vowel dimension. Unexpectedly, the subjects from both language groups demonstrated more tolerance towards /ʂa/ variability than /ɕa/, indicated by a larger syllable space for /ʂa/

[1] Presumably the vowel /a/ used in this study should be transcribed as /ɑ/, since /a/ does not exist as an open syllable in standard English according to the traditional IPA transcribing system (Rogers, 2014).

in the stimuli grid. Based on these results, it was suggested that: 1) fricative syllable processing may be context-sensitive, as there was not a fixed set of listening strategies that was applied to both native and non-native fricatives; to be specific, when it comes to non-native fricatives, frication is not necessarily the primary cue for identification; and 2) there could be a top-down effect caused by listeners' uncertainty with non-native phonetic cues; to be specific, when listeners were exposed to unfamiliar, non-native fricative syllables, they tended to label them as a phoneme that would complete a real word, or be a part of a real word, in their native language. The fact that syllable processing may be sensitive to native and non-native listening contexts potentially challenges the assumption that listeners process fricatives across languages in the same way (Wagner et al., 2006). The result that listeners showed a lexical bias in identification serves as a reminder for experimental design. To avoid possible top-down effects, one should be aware of the phonotactic constraints of the target languages. In Mcguire's study (2007b), for the stimuli grid that was created with a syllable pair that consisted of one acceptable syllable and one phonotactically "illegal" syllable, listeners tended to label the ambiguous tokens within the stimuli grid as the syllable that was acceptable in their native language. This top-down effect as a product of the native language experience may obscure the actual cue weighting strategy of fricative processing. Therefore, studies on fricative cue weighting should avoid mixing stimuli that are and are not accepted phonotactically in the listeners' native languages as far as possible.

Studies which use similar ways of manipulating stimuli and which target vowel perception, may also bring some insights to the current study, since both vowels and fricatives consist of continuous cues. Using a syllable assimilation task, which used vowel stimuli grids consisting of a duration dimension and a spectral dimension, Morrison (2008, 2009) provided a detailed picture of the developmental progress of L1-Spanish listeners' acquisition of English vowels in terms of cue weighting. It suggested an earlier developmental stage, before a duration-dependence stage, where Spanish listeners would make use of both duration and spectral cues to distinguish English /i/ and /ɪ/. The study showed that L1-Spanish listeners' with different levels of experience with English differed in their perception, and crucially that as they became more experienced with English, cue weighting became more similar to that of native listeners. This study demonstrated that the multidimensional cue manipulation can

detect minor changes in cue dependence and syllable boundaries, and show sensitivity to perception development and changes.

The current study investigates cross-language differences in the perception of fricatives and their acoustic cues. It aims to explore native and non- native fricative-initial syllable boundaries, investigate interactions between them, to reveal potential perceptual differences between native Mandarin and native Cantonese speakers in terms of English /θ/ perception. The study used a 2-alternative forced choice task with 2-dimensionally interpolated stimuli grids, each of which covers 2 of the 3 fricative-initial syllables of interest – /fɑ/, /θɑ/, and /sɑ/. The study tested a group of native Mandarin speakers, a group of native Cantonese speakers, and a small group of native English speakers (which served as a control group).

The syllable boundaries of the stimuli grid will be estimated using a logistic mixed-effect model. If the boundary appears to be perpendicular to the fricative dimension and parallel to the vowel dimension, it means that the boundary is completely fricative driven, and the vowel step change has no effect on the identification result. Alternatively, if the boundary appears to be parallel to the fricative dimension, the boundary is completely vowel driven. In reality, as the transitional information is usually considered as a secondary cue, it is hypothesized that boundaries will be closer to perpendicular to the fricative dimension, with various levels of skewness towards the vowel dimension. This hypothesis is based on the results of various studies, which have shown that the identification of fricatives are primarily frication driven, and that transitional cues moderate the skewness of the boundary (Harris, 1958; Heinz & Stevens, 1961; Jongman, 1989; Jongman et al., 2000; Wagner et al., 2006). Another factor of interest is syllable space distribution drawn by listeners from different language backgrounds. Cantonese /fɑ/-/sɑ/ space is hypothesized to be evenly distributed, since they only have two native fricatives; meanwhile, the space distribution of Mandarin and English listeners is less easy to hypothesise due to their more complex fricative inventories. Nevertheless, cross-language differences are expected to be observed in terms of syllable space distribution, and the differences should be related to the differences in fricative inventories and the spectral differences among /f s/ across languages (see Chapter 2).

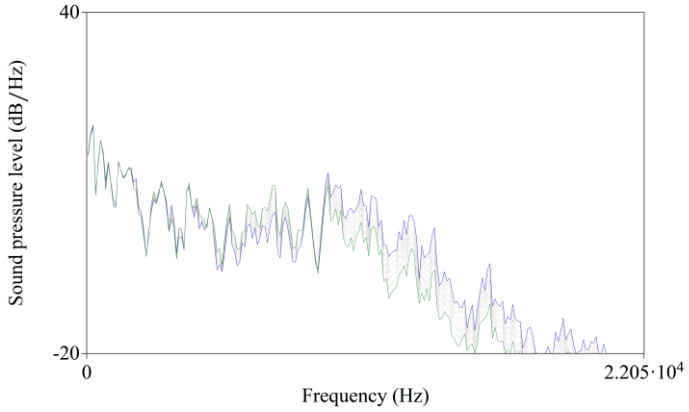
4.2 Method

4.2.1 Participants

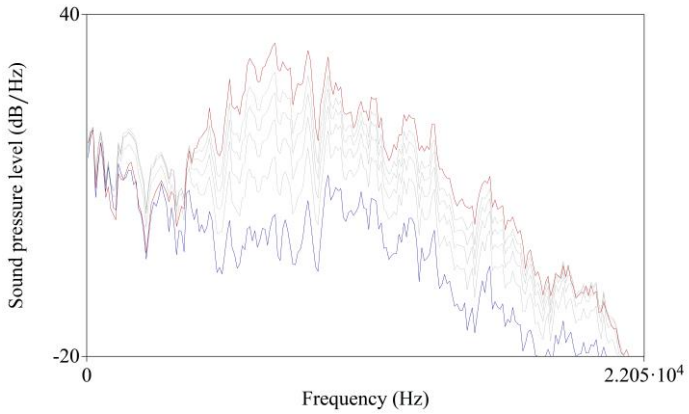
Participants were 25 native Northern Mandarin Chinese speakers (18 females and 7 males), 21 native Hong Kong / Macau Cantonese speakers (14 females and 7 males), and 10 native British English speakers (6 females and 4 males), who were all between 18 to 30 years old. All subjects reported no history of hearing, learning, or language impairment. All the native English speakers were monolinguals (were not exposed to another language before 5 years old), and had no learning experience of either Mandarin or Cantonese; the Mandarin speakers had no exposure to English before 5 years old; the Cantonese speakers had no extensive learning experience of either English nor Mandarin until 5 years old. The lab-based participants received either course credits or £4 payment on completion of the task. The web-based participants did not receive any reward (more details introduced in section 4.2.3).

4.2.2 Stimuli

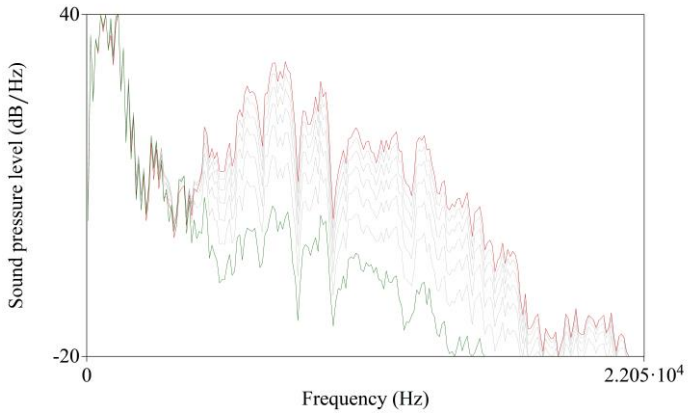
The same set of stimuli were used for both lab-based and online-based testing. A female native speaker of Standard Southern British English (SSBE) was recorded producing the syllables /fɑ/, /θɑ/, and /sɑ/, each for 5 times. She was instructed to produce them as naturally as possible. The recordings were conducted in a soundproof booth, with a Rode NT-1A microphone positioned 45 cm away from their mouths, and an RME Fireplace UC audio interface. All recordings were sampled at 44 kHz using Audacity installed on a PC. A 100 Hz high-pass Hann band filter was applied to the recordings, with smoothing frequency at 50 Hz. The average amplitude was equalised to 70 dB SPL. Three recorded syllables, which had the most similar pitch contours and no random bursts in the frication, were selected for further manipulation. These recordings were time-aligned using an overlap-add method within Praat (Boersma & Weenink, 2018), so that the initial fricative was 150 ms long and the total duration of each syllable was 550 ms. Three two-dimensional speech sound grids, with a fricative continuum and a vowel continuum as the two dimensions, were created by spectrally interpolating pairs of the recorded stimuli: /fɑ/-/θɑ/, /θɑ/-/sɑ/, and /sɑ/-/fɑ/. There were 7 steps on the fricative continuum and vowel continuum, forming a two-dimensional grid of 49 stimuli for each sound pair. The spectral interpolation was conducted using



(a) /f/ - /θ/ continuum



(b) /f/ - /s/ continuum



(c) /s/ - /θ/ continuum

Figure 4-1. Long term average spectra of the seven-step continua on the fricative dimension. The spectrum of the original /s/ is indicated by the solid red line, the spectrum of the original /f/ is indicated by the solid blue line, and the spectrum of the original /θ/ is indicated by the solid green line. The interpolated spectra are indicated by the dotted grey lines.

MATLAB with the add-on COCOHA toolbox (Wong, Hjortkjær, Ceolini, & de Cheveigné, 2018). A 33-band cochlear-scaled spectrogram was calculated separately

for the fricative portion and the vowel portion of each recorded syllable, and these spectrograms were interpolated by weighted averaging (e.g., averaging the spectrograms for /fɑ/ and /θɑ/ produced spectrograms that were 50% of each). After interpolation, each stimulus token was constructed by firstly filtering the original syllable with its original spectrogram, then filtering with an interpolated spectrogram. The long term average spectra of the fricatives are illustrated in Figure 4-1.

The stimuli were presented in 3 blocks, each of which consisted of all the stimuli in a given stimuli grid. Each stimulus was repeated 3 times. This gave a total of 441 trials for each subject (147 stimuli per grid * 3 repetitions).

4.2.3 Apparatus and procedure

This project only planned to conduct lab-based testing. However, due to the global pandemic of COVID-19, lab-based testing was seized before enough Cantonese-speaking participants were tested, and some testing was conducted online. Some measures only took place in the web-based setting to improve data quality.

4.2.3.1 Lab-based testing

Forty-seven participants completed the lab-based testing (25 native Mandarin-speaking, 10 native English-speaking, and 12 native Cantonese-speaking). All of them completed a pre-test questionnaire online at least 1 day before their test sessions, in which they reported their language experience, and confirmed that they had no history of hearing and learning impairment. When they came to their test sessions, they received an information sheet with details of the test and a consent form. They were informed that they had the right to drop out of the experiment at any point without giving a reason. During the test sessions, they completed the experiment in a quiet computer lab with a computer, with the task presented by Praat ExperimentMFC (Boersma & Weenink, 2018). They listened to the stimuli, delivered through Realtek High-Definition Audio Driver at 65 dB SPL, using Seinnheiser HD 280 over-ear headphones, and made a choice as fast as possible (as instructed) using a wired mouse.

The experiment started with a practice block including 4 randomly selected stimuli from the /f/-/s/ continuum to familiarise the subjects with the identification task. They were informed that if they had any questions regarding how to do the task, they should

ask after the practice block and before starting the next block, so that they could complete the experiment without interruptions. They were free to take short breaks in between blocks. The block orders were counterbalanced across subjects to minimize the effect of listening fatigue on the test result. The entire lab-based testing session took approximately 30 minutes.

4.2.3.2 Web-based testing

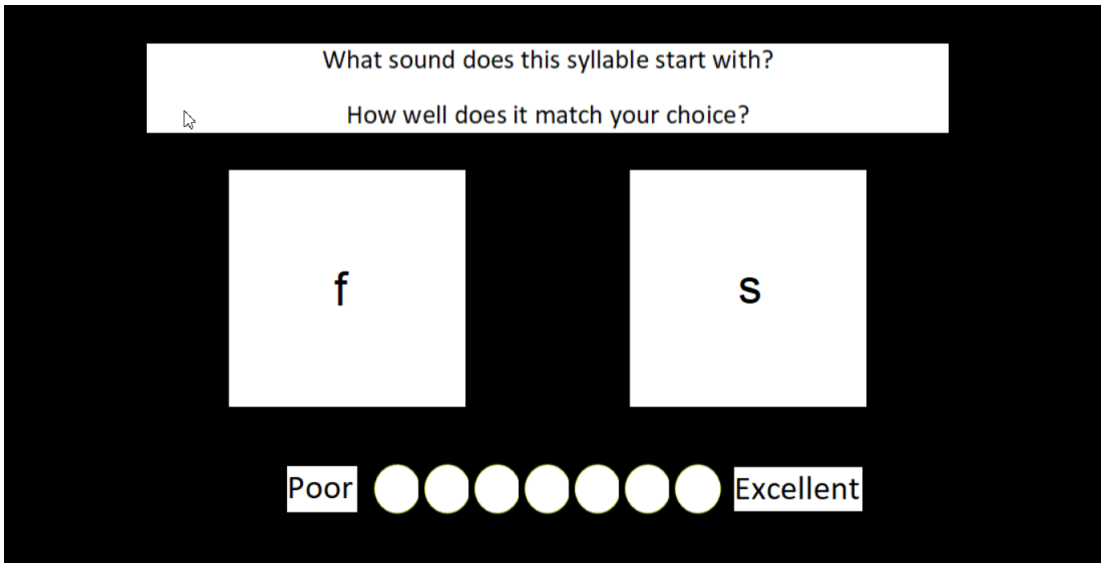
Nine native Cantonese-speaking participants completed the web-based testing, which was delivered by PsyToolkit 2.6.1 (Stoet, 2010, 2017). The test started with a detailed introduction of the task, which included the same information as the information sheet provided to the lab-based subjects. In the instructions, participants were asked to complete the test in a quiet room, uninterrupted, and with (preferably over-ear) headphones. They had the flexibility to adjust the volume according to their comfort. They then completed the same language background questionnaire as used in lab-based testing; in addition, it had a short Hong Kong Cantonese vocabulary test as a pre-screening measure, making sure that the participants were native Cantonese speakers from Hong Kong. The vocabulary list consisted 5 trendy expressions that are specific to Hong Kong Cantonese. The list was compiled based on a study by Tang (2009), and opinions of 10 native speakers of Hong Kong Cantonese and 5 native speakers of Guangdong Cantonese, making sure that the words were only comprehensible in Hong Kong Cantonese.

The structure of the task was similar to the lab-based testing. Participants first completed a practice block including 4 randomly selected stimuli from the /f/ - /s/ grid, followed by 3 experiment blocks, with stimuli of one sound pair grid in each block. The order of presentation of blocks was counterbalanced across participants. The entire web-based testing session should take about 30-40 minutes.

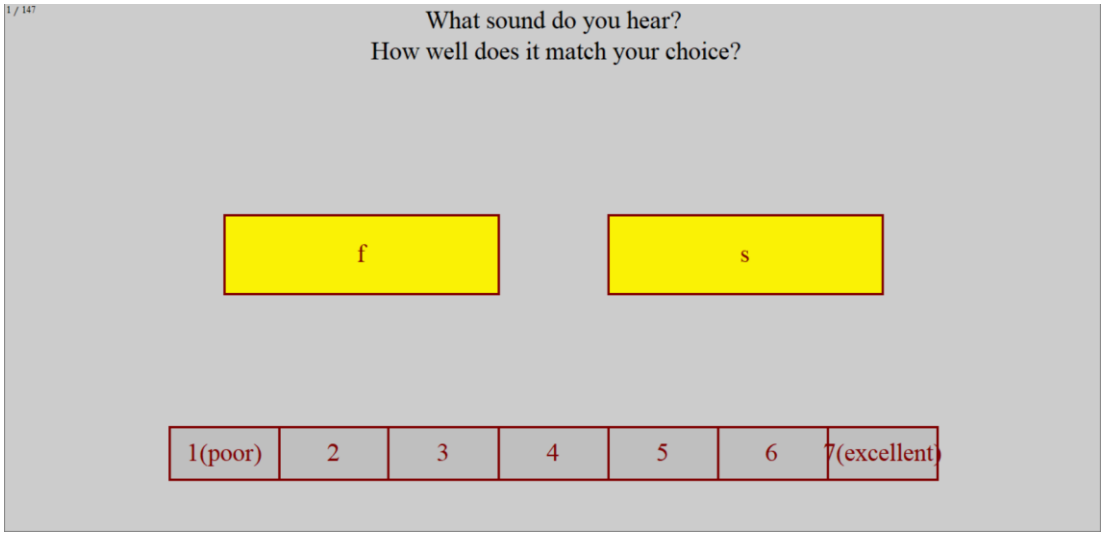
4.2.3.3 Experiment design

The same experiment design was used by the two cohorts. The experiment was a 2-alternative forced choice (2AFC) task followed by a goodness rating. On each trial, listeners heard a stimulus, categorised the initial sound and then gave a goodness rating to reflect how good an example of this sound category they thought the stimulus was. They gave their rating on a scale from 1 (poor) to 7 (excellent). The lab-based and

web-based testing interfaces are shown in Figure 4-2. The wording of the question of the web- based interface was adjusted to avoid confusion when there was no face-to-face opportunity to clarify. In each trial, a stimulus was played once. After the stimulus was presented, the participants were given 3 seconds to make their choice by clicking on one of the letters, and then 2 and 3 seconds to rate the stimulus for the lab-based and web-based participants respectively. The experiment would not proceed automatically until it recorded the goodness rating. The next stimulus was presented 1 second after the lab-based subjects rated the previous stimulus, and 1.3 seconds after the web-based subjects rated the previous stimulus.



(a) web-based testing interface



(b) lab-based testing interface

Figure 4-2. The testing interfaces of web and lab-based contexts. The subjects from both cohorts first listened to the stimuli, identified the initial fricative and used a mouse to click on the respective letter, and then to click on the rating scale.

4.2.4 Analysis

Responses in the practice block were excluded from the analysis. To exclude outliers, the responses of the participants were screened. Only participants who reached at least 50% accuracy for the fricative continua endpoints, i.e. the stimuli of fricative step 1, and fricative step 7 on each grid, in the 2AFC task were included in the analysis (cf. Yu & Lee, 2014). One English participant's /f/ - /s/ responses were excluded as its accuracy was below 50% for the /s/ end of the grid. Outlier inspection of the responses to the /f/ - /θ/ grid was exempted from this rule, as /fɑ/ and /θɑ/ are acoustically similar. If a participant reached 67% accuracy for the grid endpoint, i.e. the stimuli of fricative step 1 * vowel step 1, and the stimuli of fricative step 7 * vowel step 7, their data were included in the analysis. One Mandarin subject's /f/ - /θ/ responses were excluded in the analysis, as they were all /f/ responses for the entire block.

As each stimulus trial received a category label and a goodness rating, a score between 0 and 1 was assigned to each stimulus based on its label and the rating: 0.5 was assigned to be the category boundary, and two 7-point goodness rating scales (1-7) were mapped onto 0-0.5 (0.5 excluded) and 0.5-1 (0.5 excluded). For example, for an 'f' response with a goodness rating of 7 (excellent exemplar), it would receive a score of 1, and a 0 score for an 's' response with a rating of 7. All participants' scores were then modelled with mixed-effects logistic regression, using the *lme4* package (Bates et al., 2015) fitted in R. The model contains 1 random effect predictor, *Subject*, and 3 fixed effect predictors: *Language*, *Fricative_step* and *Vowel_step*. *Language* indexed the native language group the participants belonged to, and the 3 levels of this predictor were marked with the initial letter of each language (E for English, C for Cantonese, and M for Mandarin). Deviation coding method was used for coding the 3 levels of *Language*, so that the language groups' responses were compared individually to the grand mean of all the groups during analysis. *Fricative_step* was a 1-to-7 continuum, indexing the location of a stimulus on the fricative dimension of the stimulus grid; *vowel_step*, which was also a 1-to-7 continuum, indexed the location of a stimulus on the vowel dimension. Levy (2018) and Winter (2013) point out that the interpretation of the statistical significance of main effects with the presence of interactions requires Likelihood Ratio Tests, and so these were performed to interpret the main effect of

Language before the analysis of parameter estimates from the fitted logistic regression model.

4.3 Results

The statistical analyses were conducted in R. Unless stated otherwise, all response data from the 2AFC tasks were grouped based on stimuli grids were fitted into a generalised linear mixed-effect model with *lme4* package (version 1.1-23) fitted in R (Bates et al., 2015). Participants' native language (*Language*), the stimulus's positions on the frication dimension (*Fricative_step*), and the stimulus's position on the vowel dimension (*Vowel_step*) were the fixed factors. Participant (*Subject*) was a random factor in the models. Likelihood Ratio Tests (LRT) were conducted to investigate statistical significance of the factors and interactions following the method explained by Levy (2018) and Winter (2013), using the *anova* or *drop1* function embedded in *lme4* (Bates et al., 2015). The method compares the full model with a reduced model without the factor in question, and a fixed effect of the factor is considered significant if the difference between the likelihood of these two models is significant ($p < .05$). For each stimulus grid, there are tables showing a summary of statistics output of the predicted logistic function, including the parameter outputs (estimated coefficient, the standard error, the associated Wald's Z-score, and *p*-value) for each fix effects and interactions in the full model. There are also boxplots of performance scores, 3-D and 2-D plots illustrating estimated boundaries based on predicted functions.

4.3.1 Web-based data validation

As the setting of the web-based experiment was different from the lab-based experiment, and online testing is subject to difficulties in sustaining participants' attention and that the testing environment was less controlled, it was necessary to examine potential differences in the behavioural responses across cohorts (web-based and lab-based) before further cross-language comparisons. Figure 4-3 demonstrates all the responses while each stimuli grid condition were inspected separately. Since all the data were from the Cantonese group, *Language* was not a factor in any of the models. The main effect of, and interactions involving *Cohort* as a factor were tested using LRT. The two levels of *Cohort* were dummy coded (web-based = 0, lab-based = 1). As shown in Table 4-1, there was neither a significant main effect of *Cohort* nor

a significant interaction involving *Cohort* ($p > .05$). This means that overall performance in the experimental task was comparable across the two cohorts. As a result, all following analyses were conducted including both web and lab collected data.

Table 4-1 Summary of outputs of the Likelihood Ratio Tests inspecting the significance of main effects and interactions. Model1 is the full model that includes all the fix effects variables and their interactions, model2 is model1 excluding the *Cohort* variable, and model3 is model1 excluding the interactions.

model3: SCORE ~ cohort + fricative_step + vowel_step + (1 | subject)								
model2: SCORE ~ cohort * fricative_step * vowel_step - cohort + (1 | subject)								
model1: SCORE ~ cohort * fricative_step * vowel_step + (1 | subject)								
anova(model1, model2)								
/fɑ/-/sɑ/	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
model2	8	1535.5	1583.7	-759.74	1519.5			
model1	9	1537.7	1592.0	-759.83	1519.7	0	1	1
/fɑ/-/θɑ/								
model2	8	3666.4	3714.7	-1825.2	3650.4			
model1	9	3667.5	3721.8	-1824.7	3649.5	0.935	1	0.334
/sɑ/-/θɑ/								
model2	8	2004.4	2052.7	-994.20	1988.4			
model1	9	2005.9	2060.2	-993.94	1987.9	0.517	1	0.472
anova(model1, model3)								
/fɑ/-/sɑ/	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
model3	5	1536.7	1566.9	-763.35	1526.7			
model1	9	1537.7	1592.0	-759.83	1519.7	7.037	4	0.134
/fɑ/-/θɑ/								
model3	5	3668.5	3698.6	-1829.2	3658.5			
model1	9	3667.5	3721.8	-1824.7	3649.5	8.984	4	0.062
/sɑ/-/θɑ/								
model3	5	2002.1	2032.3	-996.04	1992.1			
model1	9	2005.9	2060.2	-993.94	1987.9	4.197	4	0.380

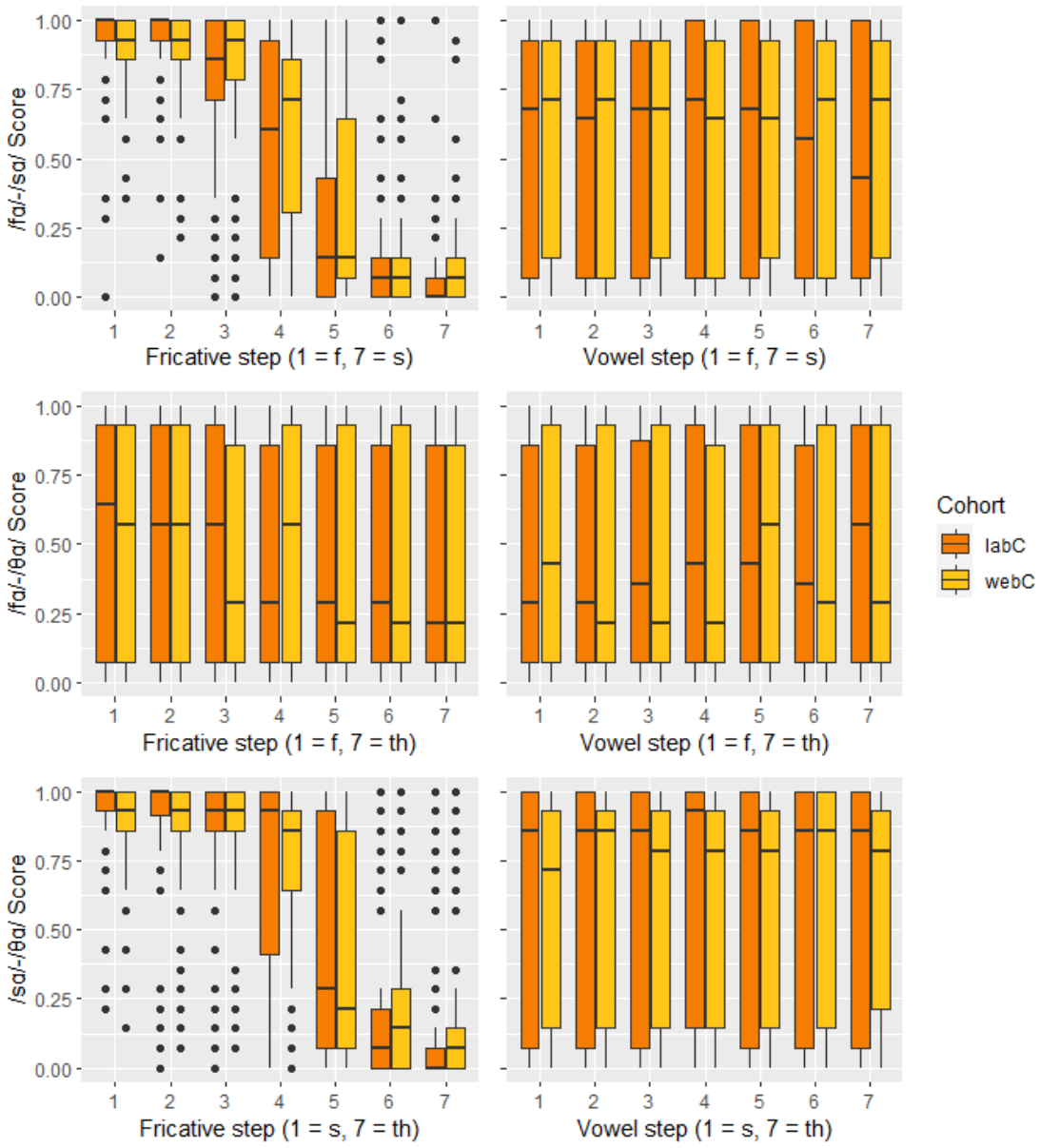


Figure 4-3 Boxplots demonstrating the calculated *Score* of all the responses of all the native Cantonese participants, divided by cohorts, stimuli grid, fricative steps, and vowel steps. LabC = lab-based group, n = 12; webC = web-based group, n = 9.

4.3.2 Cross-language comparisons

4.3.2.1 /fa/-/sa/ stimulus grid

The LRT was performed to compare two models: both included fix effects of all the independent variables, while model 2 excluded all the interactions among the independent variables which were present in model 1. The results are reported in Table 4-2, and it shows no evidence for a significant improvement of model fitting by including the interactions ($\chi^2(7) = 5.12$, $p = .6458$). The LRT was then performed again to test the significance of main effects, and the results showed significant main effects

of *Language* ($\chi^2(2) = 12.10$, $p = .003$) and *Fricative_step* ($\chi^2(1) = 6570.40$, $p < .001$), and a non-significant main effect of *Vowel_step* ($\chi^2(1) = 0.00$, $p = .876$). A Tukey's test for post hoc analysis following the LRT showed that the effect of *Language* lied in the difference between the Mandarin and the Cantonese groups. The scores of the Mandarin group (EMM = -0.03) and the scores of the Cantonese group (EMM = 0.26) differed significantly ($p = .003$).

Table 4-2 Summary of the result of Likelihood Ratio Tests that inspected the interactions and the main effects for /f/-/s/ stimuli condition.

model2: SCORE ~ language + fricative_step + vowel_step + (1 | subject)
model1: SCORE ~ language * fricative_step * vowel_step + (1 | subject)
anova(model1, model2)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
model2	6	4462.5	4504.5	-2225.2	4450.5			
model1	13	4471.4	4562.3	-2222.7	4445.4	5.116	7	0.6458

drop1(model2, test= "Chisq")

	npar	AIC	LRT	Pr(Chi)
<none>		4462.5		
Language	2	4470.6	12.1	0.003*
Fricative_step	1	11030.9	6570.4	<0.001*
Vowel_step	1	4460.5	0.0	0.876

Table 4-3 reports the parameter outputs for each fix effects in the full model. Similar to the results of the LRT, a significant fixed effect of *Fricative_step* was shown ($\beta = -1.49$, $p < .001$): when it increased, there was a decrease in the dependent variable *Score*, and it had similar impact on all language groups, as the interactions of *Language* and *Fricative_step* were not significant (Cantonese: $\beta = -0.10$, $p = .388$; Mandarin: $\beta = -0.15$, $p = .167$). One significant interaction involving *Vowel_step* and Mandarin language group was found ($\beta = 0.22$, $p = .036$), but it was not significant enough to be revealed in the overall interaction (see Table 4-2).

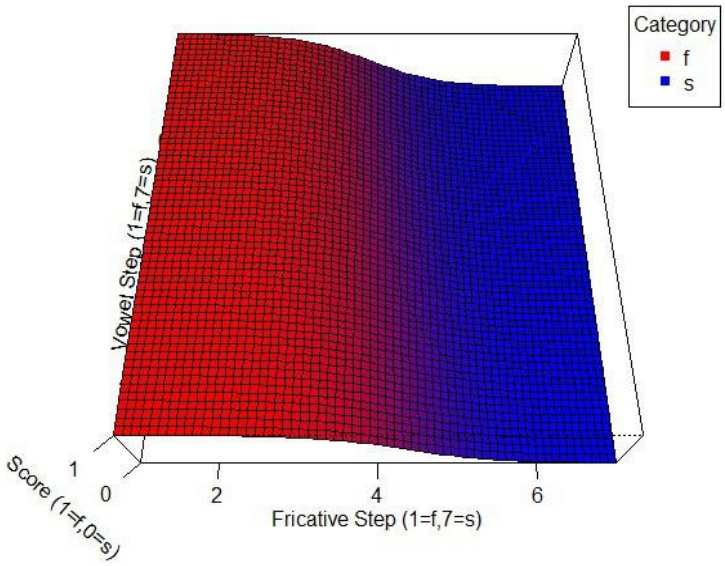
Table 4-3 Estimates for predictors in a mixed-effects logistic regression model fitting data from stimulus grid /f/-/s/. The full model formula in lme4 style was SCORE~language*fricative_step*vowel_step + (1|subject).

Predictor	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	7.644	0.464	16.478	<0.001*
LanguageC	0.652	0.525	1.242	0.214
LanguageM	-1.006	0.478	-2.105	0.035*
Fricative_step	-1.852	0.091	-20.356	<0.001*
Vowel_step	-0.155	0.084	-1.842	0.065
LanguageC:Fricative_step	-0.102	0.118	-0.864	0.388
LanguageM:Fricative_step	-0.151	0.110	-1.381	0.167
LanguageC:Vowel_step	-0.131	0.110	-1.194	0.233
LanguageM:Vowel_step	0.215	0.102	2.101	0.036*
Fricative_step:Vowel_step	0.034	0.019	1.777	0.076
LanguageC:Fricative_step:Vowel_step	0.024	0.025	0.983	0.326
LanguageM:Fricative_step:Vowel_step	-0.040	0.024	-1.716	0.086

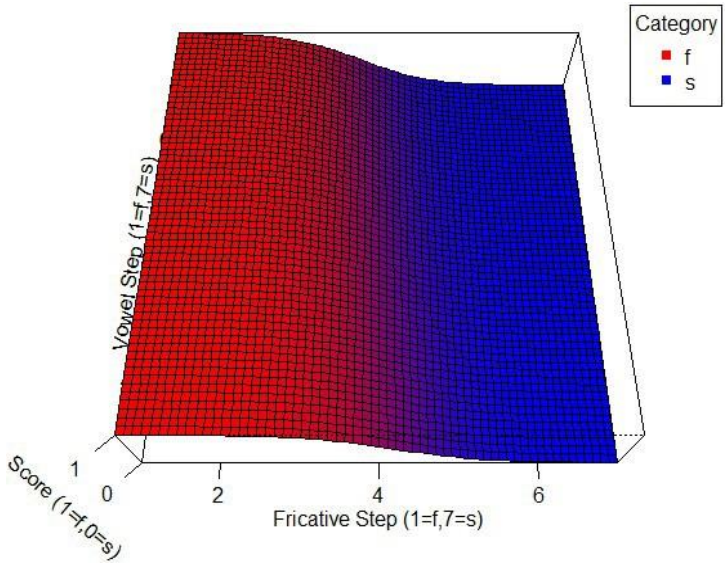
Figure 4-4 shows the predicted logistic function for each language with the estimated parameters of *Fricative_step* and *Vowel_step*. Figure 4-5(a) exhibits the distribution and boundaries of *Score* values across fricative step changes of all the language groups. The category boundaries of all three language groups appear to be roughly perpendicular to the fricative change dimension, with minor tilting in the vowel change dimension. The surface distributions of the English group and the Cantonese group appear to be similar in the current analysis, with the surface area of /f/ and /s/ roughly the same. Notably, the surface distribution of the Mandarin group appears to be uneven, showing a larger area for /s/. This means that more ambiguous stimuli (i.e. stimuli that are closer to the centre of the grid) were labelled as /s/.

Disregarding the statistically non-significant factors including *Vowel_step* and the interactions among factors, Figure 4-5(b) demonstrates the main effects of only *Language* and *Fricative_step*. The same pattern can be observed from both

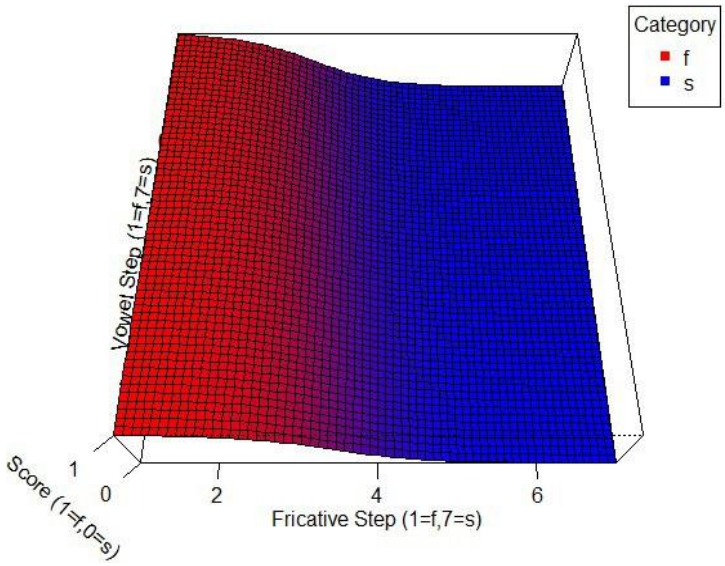
Figure *4-4* and Figure 4-5(b), as the Mandarin listeners appeared to have labelled more stimuli as /s/ than /f/, revealed by a larger space for /s/ than for /f/ in the stimuli grid when compared to other listeners, whose stimuli grid space between /f/ and /s/ appears to be evenly distributed.



(a) Cantonese

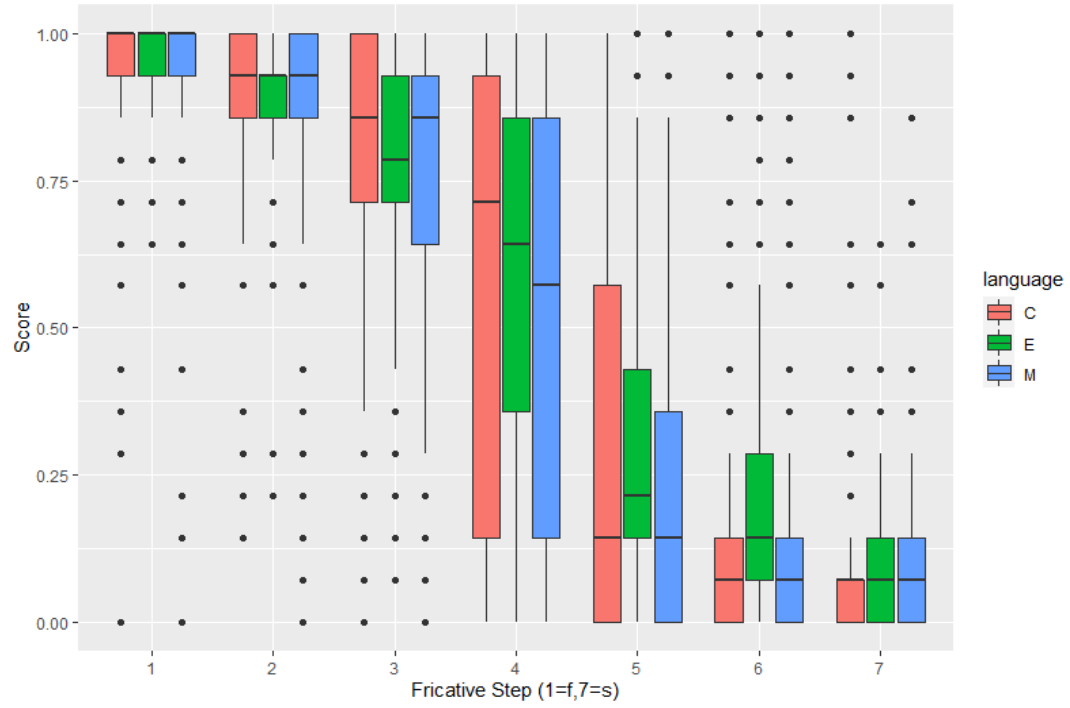


(b) English

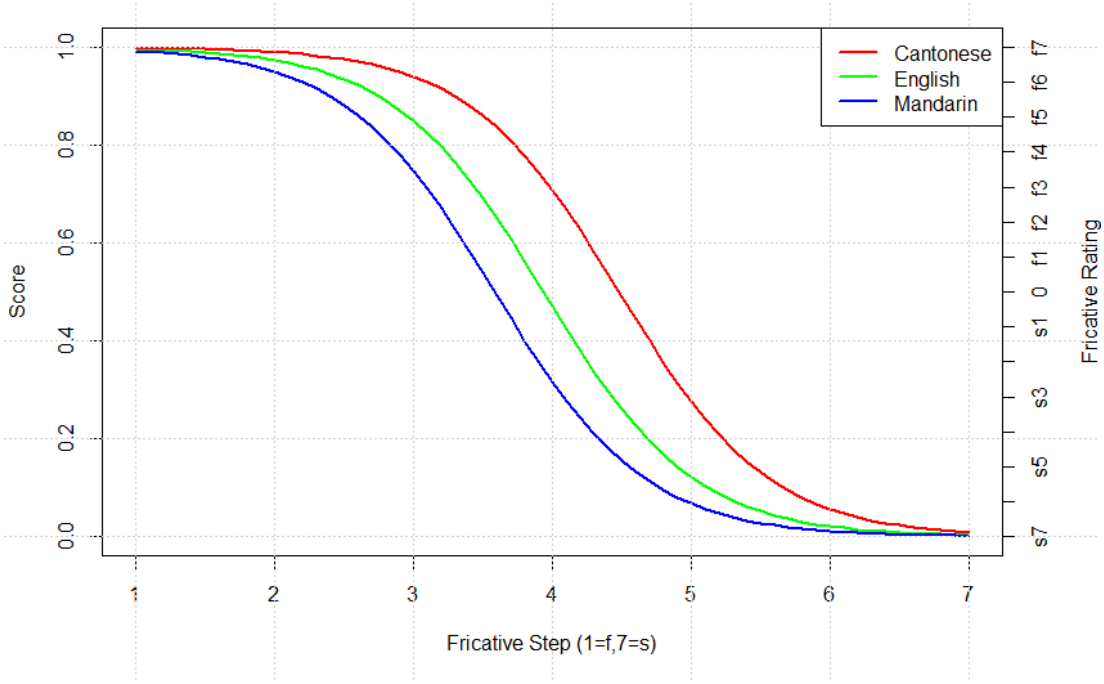


(c) Mandarin

Figure 4-4 Estimated /fɑ/-/sɑ/ boundaries of the three languages based on the estimated parameters from the mixed-effect logistic model fitted to the 2AFC task response data. The category boundaries are represented by the transition area between red and blue, which appear to be dark grey in the plots.



(a)



(b)

Figure 4-5 (a) Boxplots of calculated Score values based on raw 2AFC task response data of /fɑ/-/sɑ/ grid, excluding the non-significant factor Vowel_step; (b) 2-dimensional logistic regression plots generated based on the parameter estimates of the fitted model, presenting estimated Score as a function of the factor Fricative_step, for each language group.

4.3.2.2 /fa/-/θa/ stimulus grid

The LRT results are reported in Table 4-4. There appear to be significant interactions since including the interactions results in a significant improvement in model fitting ($\chi^2(7) = 61.85$, $p < .001$). There was a significant main effect of *Language* ($\chi^2(2) = 75.34$, $p < .001$). Post hoc comparisons using Tukey's test indicated that the significant main effect of *Language* was the result of a difference between the Cantonese group and the other 2 groups. The Cantonese group (EMM = -0.32) and the other two groups (the Mandarin group EMM = 0.09, and the English group EMM = 0.18) differed significantly ($p < .001$).

The LRT also shows a significant main effect of *Fricative_step* ($\chi^2(1) = 360.32$, $p < .001$). Table 4-5 reports the parameter outputs for each fix effects in the full model. In line with the LRT results, a significant fixed effect of *Fricative_step* was shown ($\beta = -0.25$, $p < .001$): as this variable increased, the dependent variable *Score* decreased. There was a significant interaction between *Fricative_step* and *LanguageC* ($\beta = 0.10$, $p < .001$), indicating that the Cantonese group performed differently on this dimension in comparison to the other groups. No significant main effect of *Vowel_step* and no significant interactions involving this factor was found in the analysis ($p > .05$).

Figure 4-6 shows the predicted logistic function for each language with the estimated parameters of *Fricative_step* and *Vowel_step*. Since stimuli from the /fɑ/-/θɑ/ grid were acoustically similar, the plotted surfaces appear to be different from a traditional logistic plot, and the categorical boundaries are less clear. Moreover, the estimated scores of the model were relatively low; these ranged from 0.2 and 0.8. The boundaries of all three groups still appear to be roughly perpendicular to the fricative change dimension, without obvious tilting caused by the vowel step change. The surface distributions of the English group and the Mandarin group appear to be similar, with the surface area of /f/ and /θ/ roughly the same. Notably, the predicted scores of the Cantonese group appear to be much lower than the other groups, meaning that overall they were more likely to label stimuli in this grid as /θ/.

Disregarding the statistically non-significant factor *Vowel_step* and the interactions among factors, Figure 4-7(a) exhibits the distribution of *Score* values across fricative step changes of all the language groups, and Figure 4-7(b) demonstrates the main

effects of only *Language* and *Fricative_step*. The same pattern can be observed from both Figure 4-6 and Figure 4-7(b), as the Cantonese listeners appeared to have labelled much more stimuli as /s/ rather than /f/, while the other two groups tended to label more stimuli as /f/.

Table 4-4 Summary of the result of Likelihood Ratio Tests that inspected the interactions and the main effects for /f/-/θ/ stimuli condition.

model2: SCORE ~ language + fricative_step + vowel_step + (1 | subject)
model1: SCORE ~ language * fricative_step * vowel_step + (1 | subject)
anova(model1, model2)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
model2	6	10215	10257	-5101.3	10203			
model1	13	10167	10258	-5070.4	10141	61.848	7	<0.001*

drop1(model2, test= "Chisq")

	npar	AIC	LRT	Pr(Chi)
<none>		10215		
language	2	10286	75.34	<0.001*
fricative_step	1	10573	360.32	<0.001*
vowel_step	1	10216	3.58	0.05846

Table 4-5 Estimates for predictors in a mixed-effects logistic regression model fitting data from stimulus grid /f/-/θ/. The full model formula in lme4 style was SCORE~language*fricative_step*vowel_step + (1|subject).

Predictor	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	0.859	0.235	3.650	<0.001*
LanguageC	-0.689	0.172	-4.018	<0.001*
LanguageM	0.405	0.174	2.334	0.020*
Fricative_step	-0.246	0.029	-8.387	<0.001*
Vowel_step	0.023	0.030	0.763	0.445
LanguageC:Fricative_step	0.102	0.038	2.672	<0.01*
LanguageM:Fricative_step	-0.061	0.038	-1.594	0.111
LanguageC:Vowel_step	-0.034	0.038	-0.884	0.377
LanguageM:Vowel_step	-0.019	0.039	-0.480	0.631
Fricative_step:Vowel_step	0.001	0.007	0.168	0.866
LanguageC:Fricative_step:Vowel_step	0.007	0.009	0.820	0.412
LanguageM:Fricative_step:Vowel_step	0.001	0.009	0.165	0.869

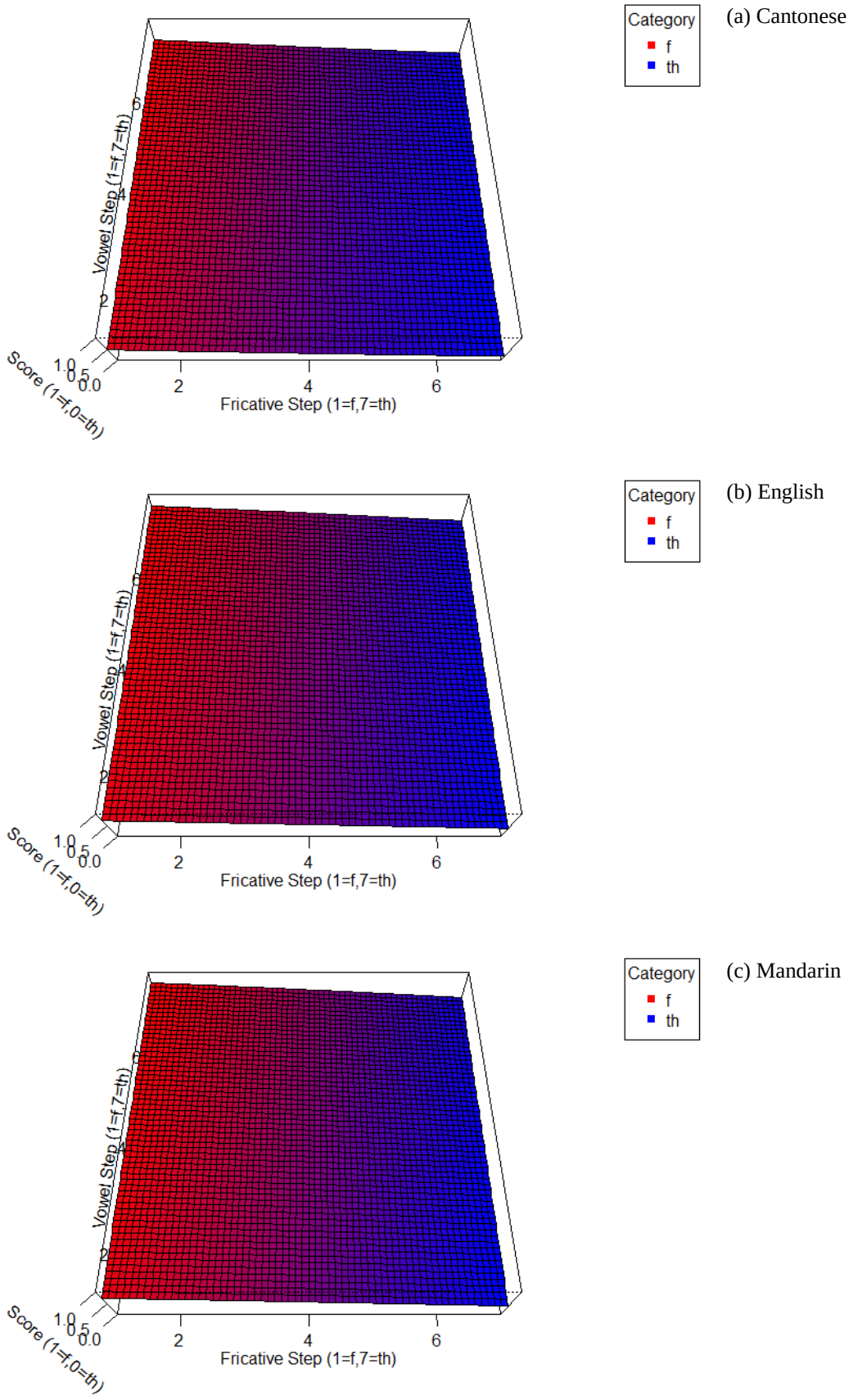
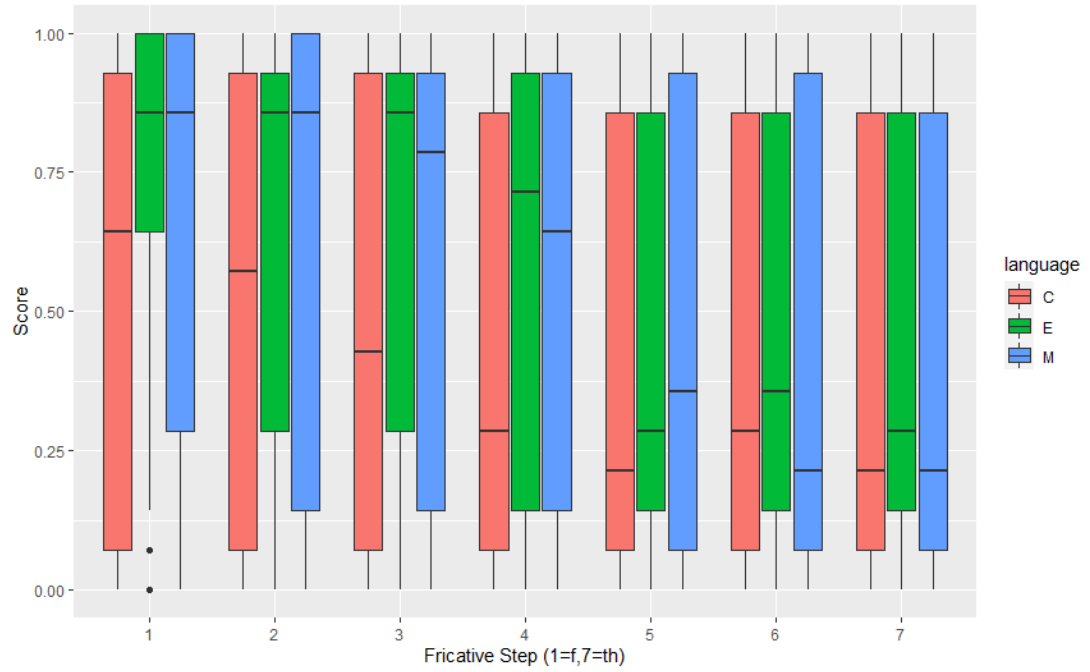
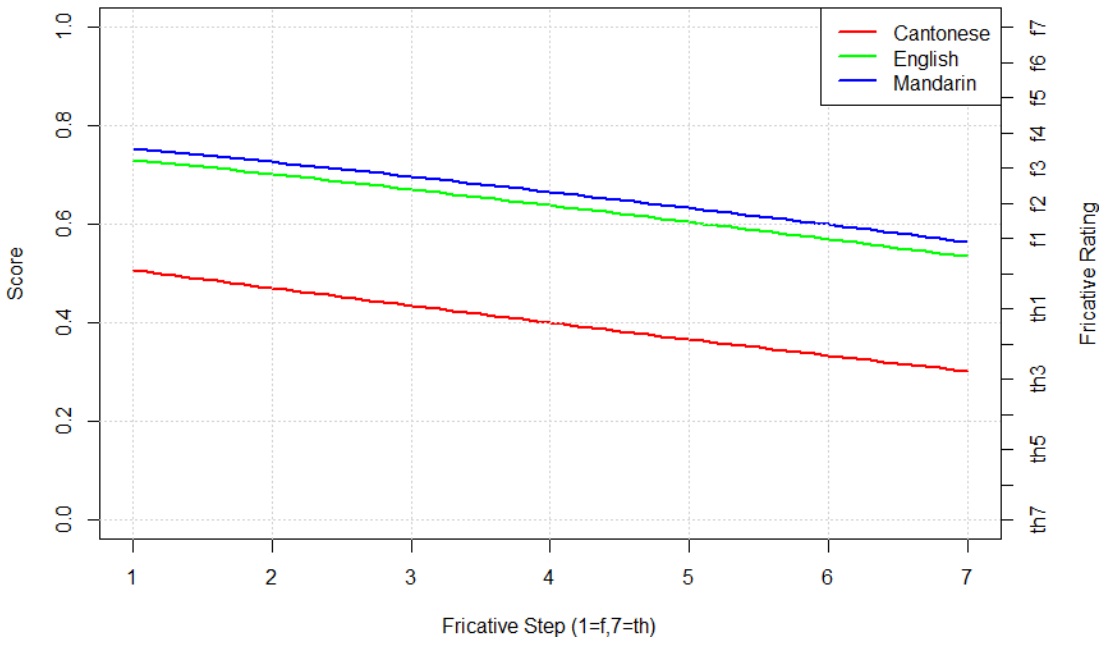


Figure 4-6. Estimated /fɑ/-/θɑ/ boundaries of the three languages based on the estimated parameters from the mixed-effect logistic model fitted to the 2AFC task response data. The category boundaries are represented by the transition area between red and blue, which appears to be dark grey in the plots.



(a)



(b)

Figure 4-7 (a) Boxplots of calculated *Score* values based on raw 2AFC task response data of /fɑ/-/θɑ/ grid, dismissing the non-significant factor *Vowel_step*; (b) 2-dimensional logistic regression plots generated based on the parameter estimates of the fitted model, presenting estimated *Score* as a function of the factor *Fricative_step*, for each language group.

4.3.2.3 /sa/-/θa/ stimulus grid

The LRT results are reported in Table 4-6. There appears to be significant interactions since including the interactions results in a significant improvement in model fitting ($\chi^2(7) = 141.93$, $p < .001$). There was also a significant main effect of *Language* ($\chi^2(2) = 42.60$, $p < .001$). Post hoc comparisons using the Tukey's test indicated that the significant main effect of *Language* was in the difference between the Mandarin group and the rest. The Mandarin group (EMM = 0.41) and the other two groups (the Cantonese group EMM = 1.01, and the English group EMM = 1.15) differed significantly ($p < .001$).

The LRT also shows a significant main effect of *Fricative_step* ($\chi^2(1) = 4412.10$, $p < .001$). Similar to the LRT result, a significant fixed effect of *Fricative_step* was shown ($\beta = -1.23$, $p < .001$, see Table 4-7): as this variable increased, the dependent variable *Score* decreased. There was also a significant interaction between *Fricative_step* and *LanguageM* was shown ($\beta = 0.35$, $p < .001$), indicating that the Mandarin group performed differently comparing to the other groups as the fricative step was changing. No significant main effect of *Vowel_step* and no significant interactions involving this factor was found in the analysis ($p > .05$).

Figure 4-8 shows the predicted logistic function for each language with the estimated parameters of *Fricative_step* and *Vowel_step*. The category boundaries of all three groups appear to be roughly perpendicular to the fricative change dimension, with very little effect of the vowel step change. Both the English group and the Cantonese group had a tendency to label more stimuli as /s/, with the surface area for /s/ slightly larger than /θ/; while the surface plot of the English group revealed a more gradual curve, showing a preference towards labelling stimuli as /s/ in this stimuli grid. Notably, the surface distribution of the Mandarin group appears to be different from the other groups, showing similar area sizes for /s/ and /θ/. This indicates that the ambiguous stimuli in this grid were labelled as /θ/ more often by the Mandarin listeners compared to the other listeners.

Disregarding the statistically non-significant factor *Vowel_step* and the interactions among factors, Figure 4-9 demonstrates the main effects of only Language and Fricative_step. The same pattern can be observed in both Figure 4-8 and Figure 4-9

(b), as the Mandarin listeners appeared to have labelled fewer stimuli as /s/ compared to the listeners from the other groups. This is revealed by an evenly distributed grid space for /s/ and /θ/, while the other listeners showed a tendency towards labelling more stimuli of this grid as /s/, with less space left for /θ/ than for /s/.

Table 4-6 Summary of the result of Likelihood Ratio Tests that inspected the interactions and the main effects for /s/-/θ/ stimuli condition.

model3: SCORE ~ language * fricative_step * vowel_step – language + (1 | subject)
model2: SCORE ~ language + fricative_step + vowel_step + (1 | subject)
model1: SCORE ~ language * fricative_step * vowel_step + (1 | subject)
anova(model1, model2)

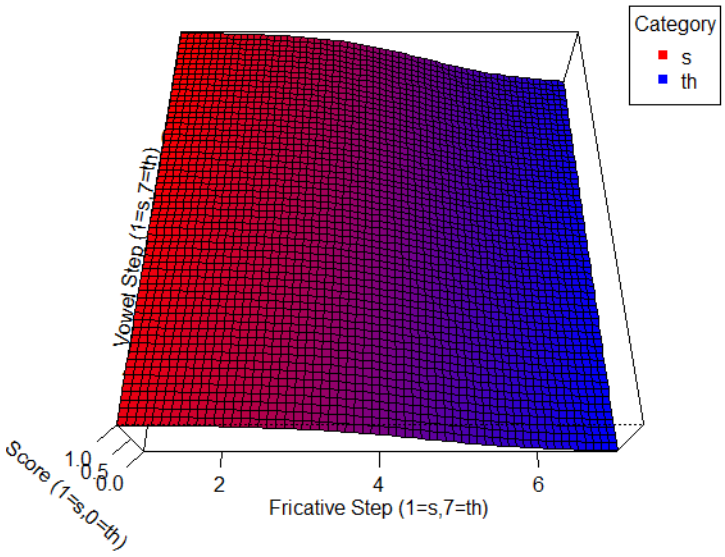
	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
model2	6	6634.5	6676.6	-3311.2	6622.5			
model1	13	6506.6	6597.8	-3240.3	6480.6	141.93	7	<0.001*

drop1(model2, test= "Chisq")

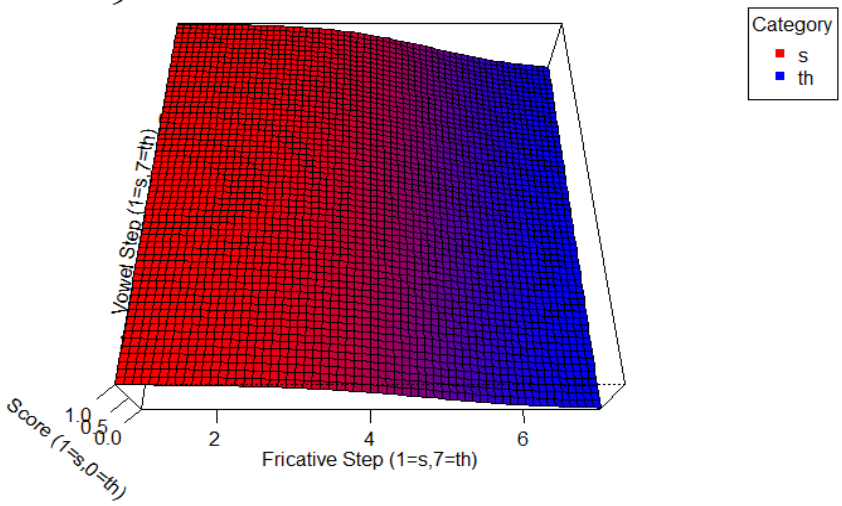
	npar	AIC	LRT	Pr(Chi)
<none>		6634.5		
language	2	6673.1	42.6	<0.001*
fricative_step	1	11044.6	4412.1	<0.001*
vowel_step	1	6632.5	0.1	0.8161

Table 4-7 Estimates for predictors in a mixed-effects logistic regression model fitting data from stimulus grid /s/-/θ/. The full model formula in lme4 style was SCORE~language*fricative_step*vowel_step + (1|subject).

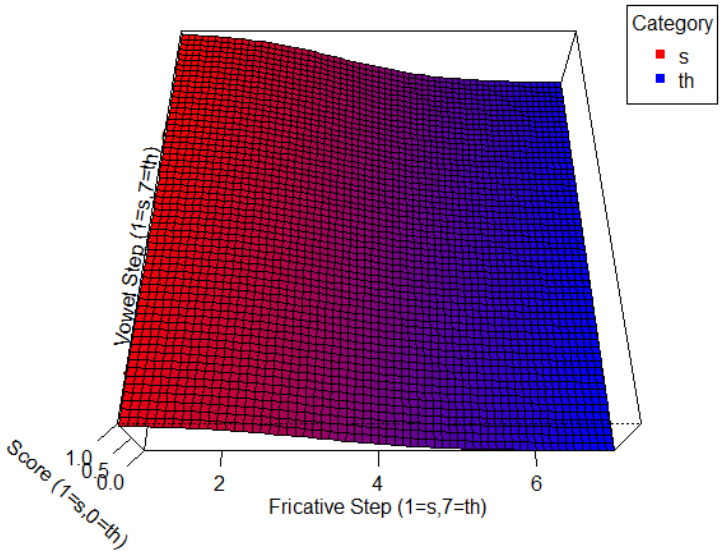
Predictor	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	5.756	0.375	15.356	<0.001*
LanguageC	-0.164	0.432	-0.379	0.705
LanguageM	-1.798	0.391	-4.592	<0.001*
Fricative_step	-1.228	0.074	-16.576	<0.001*
Vowel_step	0.032	0.081	0.396	0.692
LanguageC:Fricative_step	0.082	0.089	0.926	0.355
LanguageM:Fricative_step	0.348	0.082	4.268	<0.001*
LanguageC:Vowel_step	-0.010	0.098	-0.105	0.917
LanguageM:Vowel_step	-0.061	0.089	-0.690	0.490
Fricative_step:Vowel_step	-0.071	0.017	-0.423	0.672
LanguageC:Fricative_step:Vowel_step	0.002	0.020	0.083	0.934
LanguageM:Fricative_step:Vowel_step	0.013	0.019	0.685	0.494



(a) Cantonese

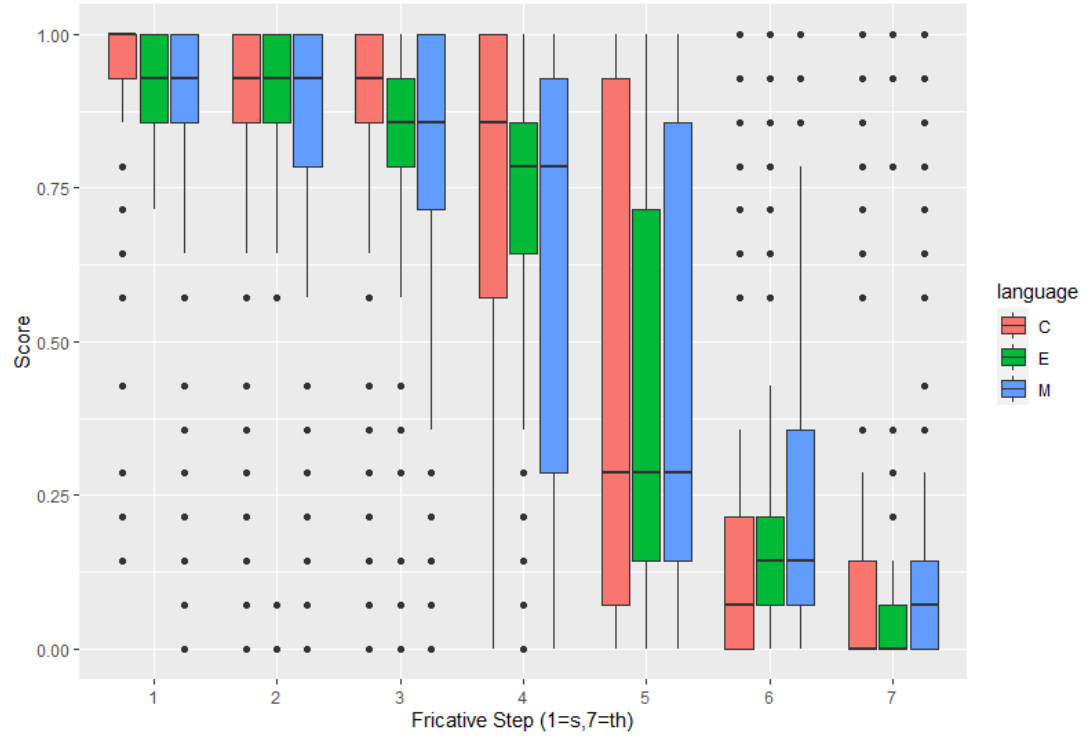


(b) English

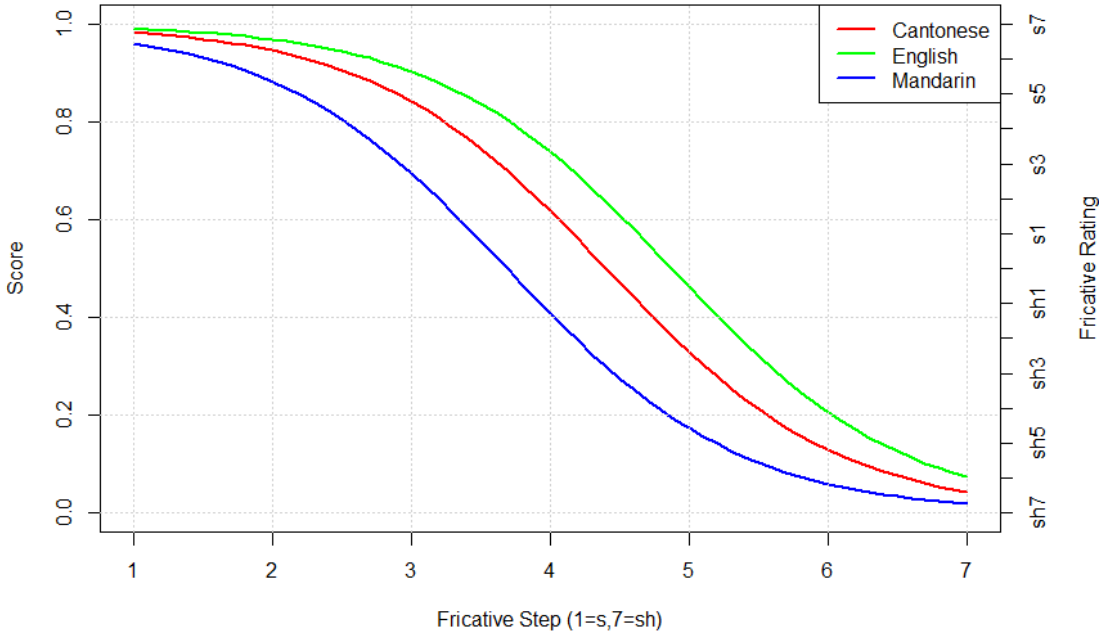


(c) Mandarin

Figure 4-8. Estimated /sɑ/-/θɑ/ boundaries of the three languages based on the estimated parameters from the mixed-effect logistic model fitted to the 2AFC task response data. The category boundaries are represented by the transition area between red and blue, which appears to be dark grey in the plots.



(a)



(b)

Figure 4-9 (a) Boxplots of calculated *Score* values based on raw 2AFC task response data of /sɑ/-/θɑ/ grid, dismissing the non-significant factor *Vowel_step*; (b) 2-dimensional logistic regression plots generated based on the parameter estimates of the fitted model, presenting estimated *Score* as a function of the factor *Fricative_step*, for each language group.

4.4 Discussion

The present study aimed to investigate possible cross-language differences in fricative syllable space distribution. Significant main effects of *Language* and *Fricative_step*

were found in all the stimuli grids, but no significant main effect of *Vowel_step* was found. The significant main effect of *Fricative_step* was predicted, as the identification and categorisation of fricatives has typically been considered to be driven primarily by the change in the spectral information included in frication (Harris, 1958; Jongman, 1989; Jongman et al., 2000; Stevens, 2000; Wagner et al., 2006; see Chapter 3 for a review). As the spectral composition proportion changed within the frication, the categorisation labels and goodness ratings of the stimuli changed accordingly; this remained the case across language groups. On the other hand, it was surprising to discover that there was no significant main effect of *Vowel_step* across language groups, even for the more acoustically similar syllable pair: /fɑ/ and /θɑ/. This means that despite the stimuli pair condition, formant transitional cues did not contribute significantly to the categorisation decisions of the listeners of any language groups.

Although there was a significant main effect of *Language* for all the stimulus grids, there were differences in how the effect was demonstrated in different grids. For the /fɑ/-/sɑ/ grid, the Mandarin group differed significantly from the Cantonese group and the English group, while the latter two groups performed similarly. As shown in Figure 4-5(b), the Mandarin group had a larger space for /sɑ/ than for /fɑ/, while Cantonese and English groups had a more evenly distributed space. For the /fɑ/-/θɑ/ grid, the Cantonese group performed differently from both the Mandarin and the English groups; Cantonese listeners categorised most of the stimuli as /θɑ/, while Mandarin and English listeners categorised most of the stimuli as /fɑ/. For the /sɑ/-/θɑ/ grid, the Mandarin group performed differently from the Cantonese and the English groups; they categorised more of the stimuli as /θɑ/. This was surprising as the spectral intensity of /s/ is usually higher than /θ/ (Jongman et al., 2000), and so it was expected that the ambiguous stimuli (i.e. stimuli closer to the centre of the grids) would have sounded closer to /s/.

The current study also aimed to investigate how listening strategies would be affected by fine-grained mismatches in formant transitions. The results of the study presented in Chapter 3 showed similar level of attention to formant transition mismatch across language groups. In contrast, in this study detailed formant transitional information did not significantly modify the syllable space distributions across language groups and stimuli grids. In other words, the listeners did not make use of the formant

transitional cues to help them categorise the stimuli in each pair, even the supposedly acoustically similar syllable pairs. The seemingly conflicting results can be explained by the fact that the two experiments involved different perceptual tasks; listeners may only attend to the cues that help them succeed at a specific task, and thus listeners in these tasks may have adopted completely different listening strategies. Compared to the phoneme monitoring task in Chapter 3 (scanning stimuli for a specific sound), the current experiment was closer to a real-life L2 learning situation in which a listener is introduced to a new L2 sound, as they have to assimilate it to a native sound category when processing the new sound (admittedly, some L2 sounds may establish a new category; however, in terms of English fricatives, they are all assimilated to native categories by Cantonese and Mandarin listeners). In this case, paying attention to too many acoustic cues may get in the way of them trying to make an assimilation decision. Based on the results of the current study, it appears that the Mandarin and Cantonese listeners deemed the transitional cues unnecessary when segregating the perceptual space of fricative syllable pairs, both native and non-native ones. We may therefore conclude that the different assimilation patterns for /θ/ for Cantonese and Mandarin native speakers is not related to different levels of dependence on coarticulatory cues.

This conclusion is in line with the study by LaRiviere et al. (1975), which pointed out that it could be the vowel instead of the coarticulation within a syllable that contributes to the differentiation between /f/ and /θ/. In a syllable identification task, they showed that the identification accuracy for /f/ was not affected by the absence of fricative-vowel transition, and was independent of the duration of the inter-syllable interval (ISI). Meanwhile, the identification accuracy of /θ/ syllables was only affected by ISI, not the absence of transition, as transitionless /θ/ syllables were well recognised when the ISI was at a certain level (such as 80 and 160 ms for /θɑ/). For both fricatives, the the identification score was largely affected by the vowel context, as /θi/ appeared to have created much more confusion than /θɑ/ and /θu/. In the current study, there was only a single vowel context /ɑ/, so it could not provide evidence to support the claim regarding the contribution of vowel context. However, the results of both studies are consistent with the conclusion that listeners did not recruit transitional cues to accomplish fricative identification. A study by Fowler (1984) provided evidence for the role of transitional cues for the perception of stops. This study suggested that acoustic cues within transition may not serve to identifying segments of a syllable, but

to isolate segments for easier identification. More discussion on the role of coarticulation in perception is in Chapter 5.

The findings also demonstrated that perception of formant cues did not vary depending on the stimuli grids; specifically, listeners did not make more use of cues from formant transitions for the more confusing stimuli pair /fɑ/-/θɑ/. However, the perception of frication cue may be dependent on the context, i.e., the fricative pair. For instance, Mandarin listeners attributed more perceptual space to /s/ in the /fɑ/-/sɑ/ grid, but gave less space to /s/ in the /sɑ/-/θɑ/ grid. It appears that spectral intensity of the frication is not the only driving factor.

The boundaries for the Mandarin and Cantonese listeners for the non-native syllable pairs potentially demonstrated perceptual compensation. Perceptual compensation refers to a phenomenon where listeners respond to speech sounds with consideration of information other than the acoustic cues, causing a reliable shift in category boundary as a function of phonetic context (such as vowel context, speech styles), and whether it is an L1 or L2 listening environment (Darcy et al., 2007; Vitela et al., 2013; Xie et al., 2017; Yu & Lee, 2014). In this experiment, listeners may have been conscious that they found it difficult to differentiate a particular syllable pair. This may have caused them to overcompensate and to have led to them being less sensitive to the changes along the spectral/fricative dimension. This in turn may have led to them labelling more ambiguous stimuli as /θ/. Indeed, compared to other listeners, Mandarin listeners were more likely to report hearing /θ/ than /s/, perhaps because they took into account their difficulty differentiating /θ/ and /s/ when labelling the stimuli. Meanwhile, Cantonese listeners were more likely to label the ambiguous stimuli as /θ/ in the /fɑ/-/θɑ/ grid, perhaps due to a similar motivation. Notably, the non-English listeners only demonstrated perceptual compensation effects in the syllable pair of /θ/ and the fricative category to which the listeners would assimilate /θ/. Namely, the Mandarin listeners only showed perceptual compensation in the /sɑ/-/θɑ/ grid, and the Cantonese listeners only in the /fɑ/-/θɑ/ grid. In contrast, the Mandarin group performed similarly to the English group for the /fɑ/-/θɑ/ grid, and the Cantonese group performed similarly to the English group for the /sɑ/-/θɑ/ grid, showing no perceptual compensation. These findings support the notion that Mandarin speakers and Cantonese speakers have different assimilation patterns for /θ/, and that this is, at

least in part, due to awareness of the listening context; to be specific, in the current study the non-native English listeners were aware of their difficulty in distinguishing /θ/ and then overcorrect by categorizing more ambiguous stimuli as /θ/. Whether listeners with limited English learning experience would have similar syllable boundaries is not clear, yet worth investigation to reveal whether this boundary shift is a cause or consequence of different assimilation patterns.

Since /fɑ/-/sɑ/ is a fricative syllable pair that is present in all 3 languages, the boundaries of the /fɑ/-/sɑ/ stimuli grid may reliably reflect native category distributions, and may account for the different assimilations of /θ/ by Mandarin and Cantonese listeners. As the Mandarin speaking participants showed a larger /s/ space in the grid, it revealed their higher level of tolerance to acoustic variability when labelling a stimulus as /s/; relatively, the Mandarin /f/ category is perceptually more restrictive. This contrast enables /s/ to be a more tolerant, and hence more suitable candidate category for the assimilation of /θ/. This argument is in line with a hypothesis of SLM-r (Flege & Bohn, 2021), which suggested that the more precisely defined an L1 category is, the more likely it can discern its difference from an L2 category. In the case of Mandarin fricatives, the /f/ category appears to be more defined, as fewer ambiguous stimuli were labelled as /f/. Instead, the Mandarin listeners showed preference towards labelling the ambiguous tokens as /s/. In comparison, /s/ performs as a less defined category in Mandarin, which allows it to assimilate /θ/. On the other hand, the Cantonese listeners showed a similar grid distribution as the English listeners. To be specific, their /fɑ/-/sɑ/ stimuli grid was evenly divided by the boundary drawn by both the Cantonese and the English listeners. In this case, assimilation of /θ/ to their /f/ category can also be explained by this similarity between the Cantonese and English listeners, considering English speakers also consider them perceptually and acoustically alike (Tabain, 1998; Wagner et al., 2006).

As /fɑ/-/sɑ/ is a fricative syllable pair that is shared by all groups, we may also attempt to connect the category boundaries to native fricative inventories. The Cantonese and English groups shared similar distribution of the /fɑ/-/sɑ/ space, despite their different native fricative inventories. As Cantonese only has /f/ and /s/, it is understandable that the division of the /fɑ/-/sɑ/ space was relatively equal. It is less clear why the

English group would also divide the space evenly, since natively they have /θ/ which exists in between /f/ and /s/. Notably, both groups of listeners tend to misperceive /θ/ as /f/ (Chan & Li, 2000; Harris, 1958; LaRiviere et al., 1975; Meng, Zee, et al., 2007), and the ambiguous tokens in the middle of the /fɑ/-/sɑ/ grid were categorised similarly across the two groups, despite their different fricative inventories. Evidently then, the presence of /θ/ in English was not reflected by the /fɑ/-/sɑ/ grid boundary drawn by the English listeners. As the Mandarin listeners had a larger space for /sɑ/ than for /fɑ/, considering their fricative inventory, one may question if the larger perceptual space for /sɑ/ is relevant to the presence of other fricatives which are spectrally similar to /s/, i.e. /ɕ/ and /ʂ/. The Mandarin /s/ has a more fronted place of articulation than English /s/, revealed by the location of its most prominent spectral peak (Lee et al., 2014), likely to be related to the existence of /ɕ/ and /ʂ/, which are crowding the articulatory space between the teeth and the hard palate. A more fronted place of articulation could not be linked directly to the larger perceptual space of /sɑ/ of the Mandarin listeners, as intuitively the former is more likely linked to less perceptual tolerance. We may tentatively conclude that the perceptual distribution of /fɑ/-/sɑ/ space could not reliably reflect the presence of /ɕ/ and /ʂ/. The Mandarin /ɕ/ and /ʂ/ were not of prior interest, thus were left out of the current study. To further investigate the link between perceptual space distribution of fricatives and their places of articulation, it is necessary to take a more holistic approach and include all the fricatives in the inventory. Generally speaking, the perceptual space boundaries demonstrated in this study did not have a reliable connection with native fricative inventories.

To conclude, the present study revealed cross-language differences in category boundaries of native and non-native fricative pairs, which could explain the different assimilation of /θ/ by Cantonese and Mandarin speakers. The findings of this study also enhanced the primary status of frication cues in fricative identification, and discovered that coarticulatory cues did not contribute much. The fact that the boundaries did not vary according to native fricative inventories indicated that one may not predict fricative boundaries, and thus L2 fricative categorisation patterns, based on information of the native fricative inventory.

Chapter 5 General Discussion

5.1 Summary of findings

Previous research had established that Mandarin listeners perceive /θ/ as /s/, and Cantonese listeners perceive /θ/ as /f/, though /f/ and /s/ exist in both languages' native fricative inventories (see Chapter 1). This thesis aimed to discover the motivation for the differential substitution of English /θ/ by Cantonese and Mandarin native speakers through 3 studies, provided more information on the acoustic properties of relevant voiceless fricatives, and explored the interactions between L1 and L2 cue weighting strategies. Study 1 (see Chapter 2), which analysed the acoustic features of the relevant fricatives, especially their spectral properties, and their efficiency in distinguishing places of articulation. The study discovered not only detailed differences among /f/ and /s/ categories across languages, but also that the smaller the native inventory, the more efficient each spectral cue was. Moreover, the attempt to model fricative processing revealed discrepancies between the assimilation patterns demonstrated by the models based only on acoustic cues and by the speakers. Study 2 (see Chapter 3), further investigated the processing of transitional cues from coarticulation, and revealed that all the listeners paid attention to the complete mismatch of fricative to vowel transition. Their fricative perception was affected, despite their different language backgrounds. Additionally, the listeners' EEG and behavioural performance was not predicted by their native fricative inventories. Study 3 (see Chapter 4), examined fricative categorisation boundaries of target fricative pairs, and it revealed significant cross-language differences driven by the spectral changes in the frication, and non-significant contribution of the coarticulatory cues. In addition, no concrete link between the fricative boundaries and the native inventories was discovered in the study; in other words, the native inventories did not predict how the fricative boundaries were drawn.

5.2 Implications

5.2.1 Different assimilations of L2 fricatives

Cross-language differences were observed in the acoustic properties of the fricatives of different languages which have the same phonemic label. The results of the present

thesis demonstrated that the differential assimilation of English /θ/ for Mandarin and Cantonese was primarily driven by the different fricative boundaries drawn for the "same" fricative across the different languages. Study 1 demonstrated that there were acoustic differences in the fricatives with the same phoneme labels, but that there was no single cue difference which corresponded to the fricative boundaries. It is therefore likely to be the combination of these differences, instead of one or two variations in specific cues, that led to the different L2 assimilation patterns as the main factor. This finding is in line with the conclusions of some other studies (e.g., Li, Munson, Edwards, Yoneyama, & Hall, 2011; Polka, 1992). This thesis has enhanced frication's primary status in both L1 and L2 fricative categorisation, despite the different language backgrounds of the listeners. Formant transition/coarticulatory cues, on the other hand, appeared to be attended to by listeners across language backgrounds during perception, but they did not contribute to fricative categorisation. This finding is largely in line with the findings of Borzone de Manrique and Massone (1981) and Zeng and Turner (1990), as these studies also concluded that, for fricative categorisation, frication cues are crucial while coarticulatory cues are not.

In the present thesis, L1 and L2 fricative categorisation boundaries could not be explained entirely by their respective fricative acoustics. It was already known that the different L2 assimilation patterns cannot be accounted for by native fricative inventories (on a phonemic level); this thesis has revealed that fricative categorisation and cue weighting strategies also could not be entirely predicted by native fricative acoustics (on a phonetic level). Similarities between L1 and L2 fricative acoustics may not predict that how these acoustic features will be perceived, and what the assimilation patterns will be. In a study by Strange, Levy and Lehnholf Jr. (2004), a similar conclusion was reached. The study discovered that front rounded vowels of French and German were assimilated to back rounded American English (AE) vowels, even though the front rounded vowels were acoustically more similar to the front unrounded AE vowels. Therefore, it appeared that acoustic similarity could not predict perceptual similarity and assimilation patterns. In fact, the PAM-L2 model has pointed out that assimilation patterns are not driven by acoustic similarities on a phonetic level, but instead by whether there is a "similar contrastive relationship" to the surrounding phoneme categories in one's phonological space (Best & Tyler, 2007, p. 28). A "similar contrastive relationship" between an L1 and an L2 sounds entails not only

similar phonetic features, but also relevant phonological rules involving the sounds in the respective language system. More importantly, it is not uncommon that perceived phonetic similarities are overridden by other phonological factors during the L2 sound assimilation process. Polka (1992) provided an example that may support this notion. In her study in which native English and Farsi listeners were tested in their perception of glottalised /k'/-/q'/ contrast in Salish (which was unfamiliar to both groups), she found that the two language groups performed comparably, even though Salish has uvular-velar place contrast and English has not. This means that having a native uvular-velar place contrast did not help Salish listeners better perceive the non-native uvular-velar contrast. Relatively speaking, the perception of a non-native phonemic contrast was not easier for listeners who had a parallel phonemic distinction in their native language than for listeners without such a distinction, likely because the former group of listeners do not prioritise this contrast during perception despite their native experience. Another piece of evidence comes from the perception of English /θ/ by native Thai listeners (Kitikanan, 2017): their assimilation of /θ/ appeared highly sensitive to vowel context, such that the assimilation patterns were sometimes at the expense of similarity in manner of articulation, e.g., /θ/ assimilated to /t/, possibly because frication features were not prioritised during the assimilation process when compared to other factors like transitional cues. One possible explanation is that native listening strategy is specific to particular combinations of phonetic features and phonological rules, and that listeners select a subset of features and rules from the whole native phonological system; that is, a fricative cue weighting strategy is not only defined by native fricative inventories or detailed frication cues, but is also influenced by other factors in the phonological system.

It is a common to investigate L2 production as a gateway to understanding L2 perceptual difficulties, but this thesis has demonstrated that acoustic similarity does not necessarily lead to perceptual similarity; in other words, fricative perception does not correspond completely to its production. To be specific, in Study 1 the perception of L1 and L2 fricatives was modelled based solely on production information, free from the influence from other phonological factors. However, this model did not predict the perceptual patterns found in listeners in Study 3. It is frequently observed that perception and production may be misaligned (e.g. Kleber, Harrington, & Reubold, 2012; Tatham & Morton, 2016). This misalignment between perception and

production also supports the notion discussed above, that the perceptual assimilation of an L2 sound is determined by many factors in a phonological system, and not just driven by phonetic similarity. This means that other than the acoustic cues from production, other phonological factors may also participate in the perception process. Consequently, it is unreasonable to assume that the perceptual assimilation may faithfully reflect the acoustic information of production, and vice versa.

5.2.2 The role of coarticulation in fricative perception

Chapter 3 revealed no significant cross-linguistic differences in listeners' sensitivity to mismatching formant transitions, indicating that the different assimilation patterns of English /θ/ were not related to the use of formant cues. This argument was further supported by findings reported in Chapter 4: evidence from perceptual experiments showed that the use of formant transition cues was not relevant to the assimilation patterns in this case. In other words, no cross-linguistic differences were found in the perception of coarticulatory cues in L1 and L2 fricative perception of Mandarin, Cantonese, and English listeners, as all of them did not depend on formant transitions to differentiate any fricative pairs. This conclusion contradicted the initial hypothesis of this study, which was established based on the findings reported by Wagner (2013) and Wagner et al. (2006). They concluded that formant transitions played a secondary role in fricative identification, and that they were more important in differentiating between perceptually similar fricative pairs such as /f/ and /θ/. This contradiction challenged the existing view of coarticulatory cues in fricative perception, and argued for a reconsideration of their role.

Regarding the role of coarticulatory cues in fricative perception, the 3 studies reported in the present thesis provided detailed information on coarticulation from various perspectives. Chapter 2 offered one measure of the transition section and how it may contribute to fricative identification. As demonstrated in section 2.3.3, the frequency of F2 at vowel onset did not reveal any cross-language differences, and it appeared similarly efficient at distinguishing places of articulation of fricatives across Cantonese, English and Mandarin. Chapter 3 provided both behavioural and EEG evidence that revealed perceptual sensitivity of all listeners to complete mismatch of formant transition when monitoring for target fricatives, despite the listeners' language

background (see section 3.3). Admittedly, the performance of the English listeners was in line with the conclusion by Wagner (2013) and Wagner et al. (2006). However, Cantonese, which is a language that only has two fricative phonemes, had demonstrated potential reliance on the transition section. On the other hand, Chapter 4 revealed that coarticulatory cues made only a limited contribution to where listeners placed boundaries within the perceptual space for fricatives.

The seemingly contradicting findings of these studies give some insights into the role of coarticulation in fricative perception. The first "conflict" is that the machine-learning-inspired imitation of cue weighting during fricative perception did not predict participants' performance in a categorisation task. In the machine-learning models, the transition section was either weighted or inhibited, but either way, was a source of information in the models. In contrast, listeners' categorisation judgements showed no influence of transition cues. One possibility is that listeners' use of coarticulatory cues may not be based entirely on what acoustic information is in the transition section and which is related to the place of articulation of the preceding fricative. To better understand this, we need to discuss the second "conflict" at the same time, which is that sensitivity to coarticulatory cues during perception did not necessarily lead to usage of those cues in fricative categorisation. This discovery is in line with the view of Fowler (1984, p. 360), who argued that "speech is segmented along coarticulatory lines into overlapping segments freed of their contextual influences". The study by Fowler (1984) conducted identification and discrimination tests which discovered that the listeners considered pairs of consonants similar when they are acoustically different but transitionally congruent, while considering pairs of consonants dissimilar when they are acoustically the same but transitionally incongruent. Thus this study concluded that coarticulatory cues were used to facilitate separation of segments from continuous speech for accurate identification of the segments. This means that listeners do pay attention to coarticulation, but that they use coarticulatory information to detect and extract the contextual influence within a segment, instead of integrating it into the perception process to identify and categorise individual segments. This explains why mismatched transitions would lead to lower identification accuracy across language backgrounds, while not contributing to category boundaries. One may conclude that, for Cantonese, Mandarin, and English listeners, coarticulatory cues were picked up in

fricative processing for isolating the fricatives from contextual influence, but that these did not participate in fricative identification.

Admittedly, although this theory fits the findings of the present thesis, it cannot account for the findings from the Dutch and German listeners in Wagner et al.'s study (2006), whose perceptual accuracy did not deteriorate despite the incongruent formant transitions. One possible explanation is that this theory does not rule out cross-linguistic differences in the perception of coarticulation. Some studies have argued that the usage of coarticulatory cues in speech perception is connected to how much coarticulation is produced, with some languages having a higher degree of coarticulation than others (M. Liang et al., 2009; Lubker & Gay, 1982; Manuel, 2009). They argue that the extent of coarticulation is generally sensitive to the phonological system of articulatory features in a language; moreover, particular features may have an impact on the perception of coarticulation. For instance, Lubker and Gay (1982) found evidence for cross-language differences in the levels of coarticulation by studying anticipatory lip rounding within syllables by Swedish and American English speakers. They found that, compared to the English speakers, the Swedish speakers produced a higher degree of coarticulation, i.e., starting to round their lips earlier on in the preceding consonant, in preparation for the following vowel. They argued that this was because Swedish vowels contrast in rounding while English vowels do not. However, lip rounding does not contribute to the identification of the consonant; instead, it may obscure the identification (e.g. Yu & Lee, 2014). Presumably, speakers who produce a higher degree of coarticulation would tend to separate the coarticulatory effects within the consonant in order to achieve accurate perception of the consonant; in other words, the coarticulatory cues may not contribute to consonant identification. Although this is a hypothesis that requires more investigation, it still points to the idea that the role of coarticulatory cues in fricative perception is not always linked to their fricative inventory. In fact, it appears to be a more complicated matter that involves more aspects of the phonological system of a language.

To summarize, the role of coarticulation in fricative perception was not relevant to the differential assimilation of /θ/ of Cantonese and Mandarin native listeners. The transition section within a CV syllable however, does contain information for place of articulation of the preceding consonant, but whether a listener makes use of this

information or not in perception is not predicted by how much acoustic information it contains, or what members there are in their native fricative inventory. While it is beyond the scope of the present thesis to establish the role of coarticulatory information in perception, possible directions for future research are revealed by the findings.

5.3 Limitations and future directions

Despite efforts to control for potential effects of language experience, this thesis cannot completely rule out effects of English learning and speaking experience on the results, since all subjects were recruited in London. The experiments controlled for participants' language experience via a language background questionnaire (see Appendix B Language Background Questionnaire), and only tested people who met the criteria set in the studies, which were; (1) having lived in an English-speaking country for less than 2 years, (2) not starting formal English training until primary school. Nevertheless, even a limited amount of exposure to an English-speaking environment may still lead changes in perception and production (e.g. Chang, 2010). Ideally, subjects of non-English speaking participants should be recruited from the places where they acquired their native language. In this way, the research findings may rule out the influence of English exposure.

Another limitation of the present thesis was due to the difficulty in matching the subject number of the Cantonese-speaking groups with the other language groups. The studies had relatively strict criteria controlling the language background of the subjects, which required the Cantonese subjects to be originally from Hong Kong, with less than 2 years of experience of living in an English-speaking country, and free from formal English training experience until primary school. The Hong Kong Cantonese in London is already a much smaller community than the Mandarin-speaking community, and due to their colonial history, a significant amount of people either have started their English learning experience before 5, or have been to an English-speaking preschool. If circumstances allow, future studies should take this difficulty into consideration, and preferably collect Cantonese data in Hong Kong.

This thesis may have revealed some task-dependent fricative processing strategies, meaning that the listeners may have adjusted their listening strategy based on the experiment task. Though the studies were designed based on an intensive review of previous research, there is a chance that these results do not faithfully reflect the listeners' fricative perception during real-life speech processing. An example was given by the study by Galle, Klein-Packard, Schreiber, and McMurray (2019): they showed that listeners were capable of making an identification decision with unnatural cues when forced to in an experiment, even when they did not do so in a natural circumstance. Thus, the findings of this study could reflect the interaction between acoustic cues and perception under certain circumstances, but they may not be generalisable to fricative processing during natural speech.

5.4 Conclusion

The present thesis investigated the motivation for cross-linguistic differences in the processing of the English fricative /θ/, and demonstrated the roles of frication and coarticulation cues in L1 and L2 fricative perception. The findings have enhanced the primary position of the acoustic cues within frication in fricative processing and identification, demonstrating that the acoustic differences among fricatives that are labelled the same by different languages appear to be the driving factor in the differential assimilation of L2 fricatives. In contrast to the view that differences in the use of coarticulatory cues are an important factor in cross-language differences in fricative perception, the present thesis has revealed no cross-language differences in the perceptual use of coarticulation, and coarticulation has a limited impact on the categorisation of both L1 and L2 fricatives. In addition, this thesis has shed more light on how the perception of L2 sounds is affected by more factors than just fricative acoustics, as well as adding to our understanding of the role of coarticulation in perception.

References

2016 Population By-census. (2016). https://www.bycensus2016.gov.hk/en/bc-mt.html

Alwan, A., Jiang, J., & Chen, W. (2011). Perception of place of articulation for plosives and fricatives in noise. *Speech Communication*, *53*(2), 195–209. https://doi.org/10.1016/j.specom.2010.09.001

Ann Burchfield, L., & Bradlow, A. R. (2014). Syllabic reduction in Mandarin and English speech. *The Journal of the Acoustical Society of America*, *135*(6), EL270–EL276. https://doi.org/10.1121/1.4874357

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bauer, R. S., & Benedict, P. K. (1997). Modern Cantonese Phonology. In *Modern Cantonese Phonology*. DE GRUYTER MOUTON. https://doi.org/10.1515/9783110823707

Behrens, S. J., & Blumstein, S. E. (1988a). On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants. *Journal of the Acoustical Society of America*, *84*(3), 861–867. https://doi.org/10.1121/1.396655

Behrens, S. J., & Blumstein, S. E. (1988b). Acoustic characteristics of English voiceless fricatives: a descriptive analysis. *Journal of Phonetics*, *16*(3), 295–298. https://doi.org/10.1016/s0095-4470(19)30504-2

Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, *January 1995*, 171–204.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.),

Second language speech learning: The role of language experience in speech perception and production (pp. 13–34). John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Boersma, P., & Weenink, D. (2018). *Praat: doing phonetics by computer [Computer program]*. Version 6.0.41. www.praat.org

Borzone de Manrique, A. M., & Massone, M. I. (1981). Acoustic analysis and perception of Spanish fricative consonants. *The Journal of the Acoustical Society of America*, *69*(4), 1145–1153. https://doi.org/10.1121/1.385694

Bybee, J. L. (2001). *Phonology and language use*. Cambridge University Press.

Chan, A. Y. W., & Li, D. C. S. (2000). English and cantonese phonology in contrast: Explaining cantonese ESL learners' english pronunciation problems. *Language, Culture and Curriculum*, *13*(1), 67–85. https://doi.org/10.1080/07908310008666590

Chang, C. B. (2010). *Title First Language Phonetic Drift During Second Language Acquisition* [University of California, Berkeley]. https://escholarship.org/uc/item/5zz4j343

Chen, H.-C., & Yip, M. C. W. (2001). Processing syllabic and sub-syllabic information in Cantonese. - PsycNET. *Journal of Psychology in Chinese Societies*, *2*(2), 199–210.

Chen, J. Y. (2000). Syllable errors from naturalistic slips of the tongue in Mandarin Chinese. *Psychologia*, *43*(1), 15–26.

Chen, J. Y., O'Séaghdha, P. G., & Chen, T. M. (2016). The primacy of abstract syllables in Chinese word production. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(5), 825–836. https://doi.org/10.1037/a0039911

Chen, T., Xu, M., Tu, J., Wang, H., & Niu, X. (2018). Relationship between Omnibus

and Post-hoc Tests: An Investigation of performance of the F test in ANOVA. *Shanghai Archives of Psychiatry*, *30*(1), 60–64. https://doi.org/10.11919/j.issn.1002-0829.218014

Cheng, C.-C. (1973). A Synchronic Phonology of Mandarin Chinese. In *A Synchronic Phonology of Mandarin Chinese*. De Gruyter Mouton. https://doi.org/10.1515/9783110866407

Cheung, H., & Chen, H. C. (2004). Early orthographic experience modifies both phonological awareness and on-line speech processing. In *Language and Cognitive Processes* (Vol. 19, Issue 1, pp. 1–28). Psychology Press Ltd. https://doi.org/10.1080/01690960344000071

Cheung, Y. S., Gan, Y. E., & Zhan, B. H. (1987). A Survey of Dialects in the Pearl River Delta. 珠江三角洲方言調查報告: 珠江三角洲方言字音對照 . In *New Century Press*. New Century Press. https://books.google.co.uk/books/about/珠江三角洲方言調查報告_珠江.html?id=cskPAAAAYAAJ&redir_esc=y

Cohen, J., & Polich, J. (1997). On the number of trials needed for P300. *International Journal of Psychophysiology*, *25*(3), 249–255. https://doi.org/10.1016/S0167-8760(96)00743-X

Connine, C. M., & Titone, D. (1996). Phoneme monitoring. *Language and Cognitive Processes*, *11*(6), 635–646. https://doi.org/10.1080/016909696387042

Cutler, A. (2000). Listening to a second language through the ears of a first. *InterpretingInterpreting. International Journal of Research and Practice in Interpreting*. https://doi.org/10.1075/intp.5.1.02cut

Cutler, A., Butterfield, S., & Williams, J. N. (1987). The perceptual integrity of syllabic onsets. *Journal of Memory and Language*, *26*(4), 406–418. https://doi.org/10.1016/0749-596X(87)90099-4

Cutler, A., & Otake, T. (1994). Mora or Phoneme? Further Evidence for Language-

Specific Listening. *Journal of Memory and Language*, *33*(6), 824–844. https://doi.org/10.1006/jmla.1994.1039

Da, J. (2010). *Chinese text computing*. Murfreesboro, TN: Department of Foreign Languages and Literatures, Middle Tennessee State University . https://lingua.mtsu.edu/chinese-computing/

Darcy, I., Peperkamp, S., & Dupoux, E. (2007). Bilinguals play by the rules: perceptual compensation for assimilation in late L2- learners. In J. Cole & J. I. Hualde (Eds.), *Papers in Laboratory Phonology 9* (pp. 411–443). Mouton de Gruyter.

Deterding, D. (2006). The pronunciation of English by speakers from China. *English World-Wide. A Journal of Varieties of EnglishEnglish World-Wide / A Journal of Varieties of EnglishEnglish World-Wide*, *27*(2), 175–198. https://doi.org/10.1075/eww.27.2.04det

Donchin, E. (1981). Surprise!...Surprise? *Psychophysiology*, *18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford University Press.

Duncan-Johnson, C. C., & Donchin, E. (1977). On Quantifying Surprise: The Variation of Event-Related Potentials With Subjective Probability. *Psychophysiology*, *14*(5), 456–467. https://doi.org/10.1111/j.1469-8986.1977.tb01312.x

Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. *Biological Psychology*, *14*(1–2), 1–52. https://doi.org/10.1016/0301-0511(82)90016-3

Eaves, M. (2011). English, Chinglish or China English?: Analysing Chinglish, Chinese English and China English. *English Today*, *27*(4), 64–70. https://doi.org/10.1017/S0266078411000563

Finney, S. A., Protopapas, A., & Eimas, P. D. (1996). Attentional allocation to

syllables in American English. *Journal of Memory and Language*, *35*(6), 893–909. https://doi.org/10.1006/jmla.1996.0046

Flege, James E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An Integrated View of Language Development: Papers in Honor of Henning Wode* (pp. 217–244). Wissenschaftlicher Verlag.

Flege, James Emil. (1995). Second Language Speech Learning: Theory, Findings, and Problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, *June*, 233–277.

Flege, James Emil, & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning* (pp. 3–83). Cambridge University Press. https://doi.org/10.1017/9781108886901.002

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America*, *84*(1), 115–123. https://doi.org/10.1121/1.396977

Fosker, T., & Thierry, G. (2004). P300 investigation of phoneme change detection in dyslexic adults. *Neuroscience Letters*, *357*(3), 171–174. https://doi.org/10.1016/j.neulet.2003.12.084

Foss, D. J., & Dowell, B. E. (1971). High-speed memory retrieval with auditorily presented stimuli. *Perception & Psychophysics*, *9*(6), 465–468. https://doi.org/10.3758/BF03208953

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, *36*(4), 359–368. https://doi.org/10.3758/BF03202790

Fowler, C. A., & Brown, J. M. (2000). Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception and Psychophysics*, *62*(1), 21–32. https://doi.org/10.3758/BF03212058

Galle, M. E., Klein-Packard, J., Schreiber, K., & McMurray, B. (2019). What Are You Waiting For? Real-Time Integration of Cues for Fricatives Suggests Encapsulated Auditory Memory. *Cognitive Science*, *43*(1), e12700. https://doi.org/10.1111/cogs.12700

Giegerich, H. J. (1992). English Phonology: An Introduction. In *English Phonology*. Cambridge University Press. https://doi.org/10.1017/cbo9781139166126

Gonsalvez, C. J., & Polich, J. (2002). P300 amplitude is determined by target-to-target interval. *Psychophysiology*, *39*(3), 388–396. https://doi.org/10.1017.S0048677201393137

Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Source: Journal of the International Phonetic Association*, *32*(2), 141–174. https://doi.org/10.1017/S0025100302001020

Hardcastle, W. J., Gibbon, F. E., & Jones, W. (1991). Visual display of tongue-palate contact: Electropalatography in the assessment and remediation of speech disorders. *International Journal of Language & Communication Disorders*, *26*(1), 41–74. https://doi.org/10.3109/13682829109011992

Harris, K. S. (1958). Cues for the Discrimination of American English Fricatives in Spoken Syllables. *Language and Speech*, *1*(1), 1–7. https://doi.org/10.1177/002383095800100101

Hashimoto, O. Y. (1972). *Studies in Yüe dialects 1: Phonology of Cantonese*. Cambridge University Press. https://books.google.co.uk/books/about/Phonology_of_Cantonese.html?id=hWf9T41DopIC&redir_esc=y

Healy, A. F., & Cutting, J. E. (1976). Units of speech perception: Phoneme and syllable. *Journal of Verbal Learning and Verbal Behavior*, *15*(1), 73–83. https://doi.org/10.1016/S0022-5371(76)90008-6

Heinz, J. M., & Stevens, K. N. (1961). On the Properties of Voiceless Fricative

Consonants. *Journal of the Acoustical Society of America*, *33*(5), 589–596. https://doi.org/10.1121/1.1908734

Hendrick, M. S., & Ohde, R. N. (1993). Effect of relative amplitude of frication on perception of place of articulation. *Journal of the Acoustical Society of America*, *94*(4), 2005–2026. https://doi.org/10.1121/1.407503

Hughes, G. W., & Halle, M. (1956). Spectral Properties of Fricative Consonants. *Journal of the Acoustical Society of America*, *28*(2), 303–310. https://doi.org/10.1121/1.1908271

Hung, T. T. N. (2000). Towards a phonology of Hong Kong English. *World Englishes*, *19*(3), 337–356. https://doi.org/10.1111/1467-971X.00183

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.2062307

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1). https://doi.org/10.1016/S0010-0277(02)00198-1

Jiang, Y. (1995). Chinglish and China English. *English Today*, *11*(1), 51–56. https://doi.org/10.1017/S0266078400008105

Johnson, K. (1991). Differential effects Of Speaker and Vowel Variability on Fricative Perception. *Language and Speech*, *34*(3), 265–279. https://doi.org/10.1177/002383099103400304

Johnson, K., & Babel, M. (2010). On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics*, *38*(1), 127–136. https://doi.org/10.1016/j.wocn.2009.11.001

Jongman, A. (1985). Duration of frication noise as a perceptual cue. *The Journal of the Acoustical Society of America*, *77*(S1), S26–S26. https://doi.org/10.1121/1.2022251

Jongman, A. (1989). Duration Of Frication Noise Required For Identification Of English Fricatives. *Journal of the Acoustical Society of America*, *85*(4), 1718–1725. https://doi.org/10.1121/1.397961

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, *108*(3), 1252. https://doi.org/10.1121/1.1288413

Katayama, J., & Polich, J. (1996). P300 from one-, two-, and three-stimulus auditory paradigms. *International Journal of Psychophysiology*, *23*(1–2), 33–40. https://doi.org/10.1016/0167-8760(96)00030-X

Kitikanan, P. (2017). The Effects of L2 Experience and Vowel Context on the Perceptual Assimilation of English Fricatives by L2 Thai Learners. *English Language Teaching*, *10*(12). https://doi.org/10.5539/elt.v10n12p72

Kleber, F., Harrington, J., & Reubold, U. (2012). The Relationship between the Perception and Production of Coarticulation during a Sound Change in Progress. *Language and Speech*, *55*(3), 383–405. https://doi.org/10.1177/0023830911422194

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*(3), 557–577. https://doi.org/10.1017/S0048577201990559

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107. https://doi.org/10.3758/BF03212211

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*. https://doi.org/10.1111/j.1467-

7687.2006.00468.x

Kühnert, B., & Nolan, F. (2009). The origin of coarticulation. In William J Hardcastle & N. Newlett (Eds.), *Coarticulation: Theory, Data and Techniques* (pp. 7–30). Cambridge University Press. https://doi.org/10.1017/cbo9780511486395.002

Kutas, M., Mccarthy, G., & Donchin, E. (1977). Augmenting Mental Chronometry: The P300 as a Measure of Stimulus Evaluation Time. In *Source: Science, New Series* (Vol. 197, Issue 4305).

Lai, L., & Cheng, S. (1991). Feature Geometry of Vowels and Co-occurrence Restrictions in Cantonese. *Proceedings of the 9th West Coast Conference on Formal Linguistics*, 107–124.

LaRiviere, C., Winitz, H., & Herriman, E. (1975). The distribution of perceptual cues in English prevocalic fricatives. *Journal of Speech and Hearing Research*, *18*(4), 613–622. https://doi.org/10.1044/jshr.1804.613

Lee, C.-Y., Zhang, Y., Li, X., Tao, L., & Bond, Z. S. (2012). Effects of speaker variability and noise on Mandarin fricative identification by native and non-native listeners. *The Journal of the Acoustical Society of America*, *132*(2), 1130–1140. https://doi.org/10.1121/1.4730883

Lee, C., Zhang, Y., & Li, X. (2014). Acoustic characteristics of voiceless fricatives in Mandarin Chinese. In *Journal of Chinese Linguistics* (Vol. 42, Issue 1). https://about.jstor.org/terms

Lee, W.-S., & Zee, E. (2010). Articulatory Characteristics of the Coronal Stop, Affricate, and Fricative in Cantonese. *Journal of Chinese Linguistics*, *38*(2), 336–372. https://www.jstor.org/stable/23754137?seq=1#metadata_info_tab_contents

Lee, W. S. (1999). *A phonetic study of the speech of the Cantonese-speaking children in Hong Kong - CityU Scholars | A Research Hub of Excellence* [City University of Hong Kong]. https://scholars.cityu.edu.hk/en/theses/theses(a4d08f83-aadb-45eb-b3af-9b1132420b7a).html

Lenth, R. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (1.5.1). https://cran.r-project.org/package=emmeans

Leung, M. T., Law, S. P., & Fung, S. Y. (2004). Type and token frequencies of phonological units in Hong Kong Cantonese. *Behavior Research Methods, Instruments, and Computers*, *36*(3), 500–505. https://doi.org/10.3758/BF03195596

Levy, R. (2018). *Using R formulae to test for main effects in the presence of higher-order interactions*.

Li, F., Edwards, J., & Beckman, M. (2007). SPECTRAL MEASURES FOR SIBILANT FRICATIVES OF ENGLISH, JAPANESE, AND MANDARIN CHINESE. *International Congress of Phonetic Sciences*, 917–920.

Li, F., Munson, B., Edwards, J., Yoneyama, K., & Hall, K. (2011). Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: Implications for cross-language differences in speech-sound development. *The Journal of the Acoustical Society of America*, *129*(2), 999–1011. https://doi.org/10.1121/1.3518716

Li, Y. (2018). The Production of Mandarin Voiceless Sibilant Fricatives by Late Cantonese- Mandarin Bilinguals : an Acoustic Study. *English Literature and Language Review*, *4*(5), 80–87.

Liang, E. (2014). Pronunciation of English Consonants , Vowels and Diphthongs of Mandarin- Chinese Speakers. *Studies in Literature and Language*, *8*(1), 62–65. https://doi.org/10.3968/j.sll.1923156320140801.4012

Liang, M., Pascal, P., & Jianwu, D. (2009). A study of anticipatory coarticulation for French speakers and for Mandarin Chinese speakers. *Chinese Journal of Phonetics*, *2*, 82–89. https://doi.org/hal-00368711

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal*

of Experimental Psychology, *54*(5), 358–368. https://doi.org/10.1037/h0044417

Liu, T., & Hsiao, J. H. (2014). Holistic Processing in Speech Perception: Experts' and Novices' Processing of Isolated Cantonese Syllables Holistic Processing in Speech Perception: Experts' and Novices' Processing of Isolated Cantonese Syllables. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*(36), 869–874. https://escholarship.org/uc/item/6sv7j3gs

Lu, Y. A. (2014). Mandarin fricatives redux: The psychological reality of phonological representations. *Journal of East Asian Linguistics*, *23*(1), 43–69. https://doi.org/10.1007/s10831-013-9111-5

Lubker, J., & Gay, T. (1982). Anticipatory labial coarticulation: Experimental, biological, and linguistic variables. *Journal of the Acoustical Society of America*, *71*(2), 437–448. https://doi.org/10.1121/1.387447

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique* (Second edi). The MIT Press. http://web.b.ebscohost.com/ehost/ebookviewer/ebook/bmxlYmtfXzc5ODM0NF9fQU41?sid=aac00a88-0a07-4704-8197-3b771fb7204c@pdc-v-sessmgr02&vid=0&format=EB&rid=1

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America*, *125*(6), 3962–3973. https://doi.org/10.1121/1.2990715

Manuel, S. (2009). Cross-language studies: relating language-particular coarticulation patterns to other language-particular facts. In William J. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, Data and Techniques* (pp. 179–198). Cambridge University Press. https://doi.org/10.1017/CBO9780511486395.009

Mcguire, G. L. (2007a). English listeners' perception of Polish alveopalatal and retroflex voiceless sibilants: A pilot study. In *UC Berkeley PhonLab Annual Report* (Vol. 3, Issue 3).

Mcguire, G. L. (2007b). *Phonetic Category Learning (Unpublished doctoral dissertation)*.

McMurray, B., & Jongman, A. (2011). What Information Is Necessary for Speech Categorization? Harnessing Variability in the Speech Signal by Integrating Cues Computed Relative to Expectations. *Psychological Review*, *118*(2), 219–246. https://doi.org/10.1037/a0022325

McNeill, D., & Lindig, K. (1973). The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 419–430. https://doi.org/10.1016/S0022-5371(73)80020-9

Meng, H., Lo, Y. Y., Wang, L., & Lau, W. Y. (2007). Deriving salient learners' mispronunciations from cross-language phonological comparisons. *2007 IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2007, Proceedings*, 437–442. https://doi.org/10.1109/asru.2007.4430152

Meng, H., Zee, E., Lee, W. S., & Lee, W.-S. (2007). *A Contrastive Phonetic Study between Cantonese and English To Predict Salient Mispronunciations by Cantonese Learners of English*.

Morrison, G. S. (2008). L1-Spanish Speakers' Acquisition of the English /i /-/I/ Contrast: Duration-based Perception is Not the Initial Developmental Stage Language and Speech. *LANGUAGE AND SPEECH*, *51*(4), 285–315. https://doi.org/10.1177/0023830908099067

Morrison, G. S. (2009). L1-Spanish Speakers' Acquisition of the English /i/—/ / Contrast II: Perception of Vowel Inherent Spectral Change1. *Language and Speech*, *52*(4), 437–462. https://doi.org/10.1177/0023830909336583

Newman, R. L., Connolly, J. F., Service, E., & Mcivor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. *Psychophysiology*, *40*(4), 640–647. http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=1&sid=20ab562d-5d6d-42aa-84a2-fc460a42b185%40pdc-v-sessmgr03

Nittrouer, S., & Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research*, *30*(3), 319–329. https://doi.org/10.1044/jshr.3003.319

Nittrouer, Susan. (1986). Acoustic consequences of fricative-vowel anticipatory coarticulation in young children. *The Journal of the Acoustical Society of America*, *79*(S1), S54–S54. https://doi.org/10.1121/1.2023280

Nittrouer, Susan, & Miller, M. E. (1997). Developmental weighting shifts for noise components of fricative-vowel syllables. *The Journal of the Acoustical Society of America*, *102*(1), 572–580. https://doi.org/10.1121/1.419730

Pallier, C., Sebastian-Gallés, N., Felguera, T., Christophe, A., & Mehler, J. (1993). Attentional Allocation within the Syllabic Structure of Spoken Words. *Journal of Memory and Language*, *32*(3), 373–389. https://doi.org/10.1006/jmla.1993.1020

Peng, L., & Setter, J. (2000). The emergence of systematicity in the English pronunciations of two Cantonese-speaking adults in Hong Kong. *English World-Wide*, *21*(1), 81–108. https://doi.org/10.1075/eww.21.1.05pen

Picard, M. (2002). The differential substitution of english /θ ð/ in french: The case against underspecification in L2 phonology. *Lingvisticae Investigationes*, *25*(1), 87–96. https://doi.org/10.1075/li.25.1.07pic

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. In *Clinical Neurophysiology* (Vol. 118, Issue 10, pp. 2128–2148). NIH Public Access. https://doi.org/10.1016/j.clinph.2007.04.019

Polich, J. (2012). Neuropsychology of P300. In E. S. Kappenman & S. J. Luck (Eds.), *The Oxford Handbook of Event-Related Potential Components* (pp. 159–188). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0089

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological Psychology*, *41*(2), 103–146.

https://doi.org/10.1016/0301-0511(95)05130-9

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics*, *52*(1), 37–52. https://doi.org/10.3758/BF03206758

Proctor, M., Hsuan Lu, L., Zhu, Y., Goldstein, L., & Narayanan, S. (2012). Articulation of Mandarin Sibilants: a multi-plane realtime MRI study. *Proceedings of the 14th Australasian International Conference on Speech Science and Technology*, 113–116. http://sail.usc.edu/span

Pulleyblank, E. G. (1997). The Cantonese vowel system in historical perspective. In *Studies in Chinese Phonology*. DE GRUYTER. https://doi.org/10.1515/9783110822014.185

Ramsey, S. R. (1989). *The Languages of China*. Princeton University Press. https://books.google.co.uk/books/about/The_Languages_of_China.html?id=2E_5nR0SoXoC&redir_esc=y

Rogers, H. (2014). The Sounds of Language: An Introduction to Phonetics. In *The Sounds of Language*. Routledge. https://doi.org/10.4324/9781315838731

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*(JUNE), 31. https://doi.org/10.3389/fnsys.2013.00031

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*, *47*(sup2), S99–S105. https://doi.org/10.1080/14992020802301167

Shadle, Christine H. (1990). Articulatory-Acoustic Relationships in Fricative Consonants. In *Speech Production and Speech Modelling* (pp. 187–209). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_8

Shadle, Christine Helen. (1985). *The acoustics of fricative consonants* [Massachusetts Institute of Technology]. https://doi.org/10.1121/1.393552

Squires, K. C., Wickens, C., & Squires, N. K. (1976). The Effect of Stimulus Sequence on the Waveform of the Cortical Event-Related Potential. In *New Series* (Vol. 193, Issue 4258).

Stevens, K. N. (2000). *Acoustic Phonetics*. The MIT Press. citeulike-article-id:12907291

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Stokes, S. F., & Fang, Z. (1998). An electropalatographic description of Putonghua fricatives and affricates. *Asia Pacific Journal of Speech, Language and Hearing*, *3*(2), 69–78. https://doi.org/10.1179/136132898805577232

Strand, E. A., & Johnson, K. (1996). Gradient and Visual Speaker Normalization in the Perception of Fricatives. *KONVENS*, 14–26. https://www.researchgate.net/publication/221565064

Strange, W., Levy, E., & Robert, L. J. (2004). Perceptual assimilation of French and German vowels by American English monolinguals: Acoustic similarity does not predict perceptual similarity. *The Journal of the Acoustical Society of America*, *115*(5), 2606.

Strevens, P. (1960). Spectra of Fricative Noise in Human Speech. *Language and Speech*, *3*(1), 32–49. https://doi.org/10.1177/002383096000300105

Sussman, H. M., McCaffrey, H. A., & Matthews, S. A. (1991). An investigation of

locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America*, *90*(3), 1309–1325. https://doi.org/10.1121/1.401923

Sussman, H. M., & Shore, J. (1996). Locus equations as phonetic descriptors of consonantal place of articulation. *Perception and Psychophysics*, *58*(6), 936–946. https://doi.org/10.3758/BF03205495

Svantesson, O., & Shi, Y. (1986). Acoustic Analysis of Chinese Fricatives and Affricates. *Journal of Chinese Linguistics*, *14*(1), 53–70. https://www.jstor.org/stable/23754218?seq=1#metadata_info_tab_contents

Tabain, M. (1998). *Non-Sibilant Fricatives in English: Spectral Information above 10 kHz*. www.karger.comhttp://biomednet.com/kargerPhonetica1998;55:107-130

Tang, S.-W. (2009). 香港「潮語」構詞的初探 [Word Formation of Hong Kong Trendy Expressions]. *《中國語文研究》 [Studies in Chinese Linguistics]*, *2*, 11–21.

Tatham, M., & Morton, K. (2016). Speech production and perception. In *Speech Production and Perception*. Palgrave Macmillan. https://doi.org/10.1057/9780230513969

Toscano, J. C., McMurray, B., & Dennhardt, J. (2010). Continuous Perception and Graded Categorization: Electrophysiological Evidence for a Linear Relationship Between the Acoustic Signal and Perceptual Encoding of Speech. In *Luck Source: Psychological Science* (Vol. 21, Issue 10).

Tyler, M., Tyler, M. D., Clot, E., Villain--Bailly, M.-S., & Pattamadilok, C. (2019). *Perceptual assimilation of English dental fricatives by native speakers of European French*. 2580–2584. https://hal.archives-ouvertes.fr/hal-02494288

Vitela, A. D., Warner, N., & Lotto, A. J. (2013). Perceptual compensation for differences in speaking style. *Frontiers in Psychology*, *4*(JUL), 399.

https://doi.org/10.3389/fpsyg.2013.00399

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological Evidence for a Postperceptual Locus of Suppression during the Attentional Blink. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(6), 1656–1674. https://doi.org/10.1037/0096-1523.24.6.1656

Wagner, A. (2013). Cross-language similarities and differences in the uptake of place information. *The Journal of the Acoustical Society of America*, *133*(6), 4256–4267. https://doi.org/10.1121/1.4802904

Wagner, A., & Ernestus, M. (2004). Language-specific relevance of formant transitions for fricative. *The Journal of the Acoustical Society of America*, *115*(5), 2392–2392. https://doi.org/10.1121/1.4780540

Wagner, A., Ernestus, M., & Cutler, A. (2006). Formant transitions in fricative identification: The role of native fricative inventory. *The Journal of the Acoustical Society of America*, *120*(4), 2267–2277. https://doi.org/10.1121/1.2335422

Wang, C., Zhang, J., & Xu, Y. (2018). Compressibility of segment duration in English and Chinese. *Proceedings of the International Conference on Speech Prosody*, *2018-June*, 651–655. https://doi.org/10.21437/SpeechProsody.2018-132

Wang, Y., Behne, D. M., & Jiang, H. (2008). Linguistic experience and audio-visual perception of non-native fricatives. *The Journal of the Acoustical Society of America*, *124*(3), 1716–1726. https://doi.org/10.1121/1.2956483

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.2141003

Wei, R., & su, J. (2012). The statistics of English in China: An analysis of the best available data from government sources. *English Today*, *28*(3), 10–14. https://doi.org/10.1017/S0266078412000235

Weinberger, S. H. (1997). Minimal segments in second language phonology. In A. James & J. Leather (Eds.), *Second-Language Speech: Structure and Process* (pp. 263–312). De Gruyter Mouton. https://doi.org/10.1515/9783110882933.263

Wilde, L. (1993). Inferring articulatory movements from acoustic properties at fricative-vowel boundaries. *The Journal of the Acoustical Society of America*, *94*(3), 1881–1881. https://doi.org/10.1121/1.407575

Wilde, L. F. (1995). *Analysis and Synthesis of Fricative Consonants*. Massachusetts Institute of Technology.

Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications*. http://arxiv.org/abs/1308.5499

Wong, A. W. K., Huang, J., & Chen, H. C. (2012). Phonological Units in Spoken Word Production: Insights from Cantonese. *PLoS ONE*, *7*(11). https://doi.org/10.1371/journal.pone.0048776

Wong, A. W. K., Wang, J., Wong, S. S., & Chen, H. C. (2018). Syllable retrieval precedes sub-syllabic encoding in Cantonese spoken word production. *PLoS ONE*, *13*(11). https://doi.org/10.1371/journal.pone.0207617

Wong, D. D. E., Hjortkjær, J., Ceolini, E., & de Cheveigné, A. (2018). *COCOHA Matlab Toolbox (Version 0.5.0). Zenodo*. https://doi.org/10.5281/zenodo.1198430

Xie, X., Theodore, R. M., & Myers, E. B. (2017). More than a Boundary Shift: Perceptual Adaptation to Foreign-Accented Speech Reshapes the Internal Structure of Phonetic Categories HHS Public Access. *J Exp Psychol Hum Percept Perform*, *43*(1), 206–217. https://doi.org/10.1037/xhp0000285

Xu, Y. (1989). Syllables and junctures. In Z. Wu & M. Lin (Eds.), *A Course in Experimental Phonetics* (pp. 193–220). Higher Education Press. http://www.homepages.ucl.ac.uk/~uclyyix/yispapers/Xu1989.pdf

Xu, Y., & Wang, M. (2009). Organizing syllables into groups-Evidence from F0 and duration patterns in Mandarin. *Journal of Phonetics*, *37*(4), 502–520. https://doi.org/10.1016/j.wocn.2009.08.003

Yu, A. C. L. (2016). Vowel-dependent variation in Cantonese /s/ from an individual-difference perspective. *The Journal of the Acoustical Society of America*, *139*. https://doi.org/10.1121/1.4944992

Yu, A. C. L., & Lee, H. (2014). The stability of perceptual compensation for coarticulation within and across individuals: A cross-validation study. *The Journal of the Acoustical Society of America*, *136*(1), 382–388. https://doi.org/10.1121/1.4883380

Zee, E., & Xu, Y. (1999). Change and Variation in the Syllable-initial and Syllable-final Consonants in Hong Kong Cantonese. *Journal of Chinese Lingsuistics*, *27*(1), 120–167. https://www.jstor.org/stable/23756746?seq=1#metadata_info_tab_contents

Zeng, F. G., & Turner, C. W. (1990). Recognition of voiceless fricatives by normal and hearing-impaired subjects. *Journal of Speech and Hearing Research*, *33*(3), 440–449. https://doi.org/10.1044/jshr.3303.440

Zhang, J., Lü, S., & Qi, S. (1982). A Cluster Analysis of the Perceptual Features of Chinese Speech Sounds. *Journal of Chinese Linguistics*, *10*(2), 189–206. https://www.jstor.org/stable/23767011?seq=1#metadata_info_tab_contents

Zheng, Y., & Iverson, P. (2016). Assimilation of English /θ/ by L1 Mandarin and Cantonese speakers. *New Sounds 2016: 8th International Symposium on the Acquisition of Second Language Speech*, 213.

Appendix A Stories created for Study 1

Cantonese story:

有個叫阿莎嘅三歲嘅女仔，佢鐘意玩，鐘意發夢，最鐘意食。有一日，佢突然之間話要出去玩，背起書包，閂埋門，就咁瀟灑地出發。係後花園曬衫嘅阿媽見佢要走，大聲嗌住佢：“你快 D 返黎喔！就黎食飯啦！”阿莎聽話有嘢食，又即刻行返屋企。阿媽笑佢貪食，但都諗住炒多碟花生俾佢食。阿莎邊生果邊睇電視，已經完全唔記得咗要出去玩嘅事。

Mandarin story:

从前，在一个小村庄里有三个平凡的木匠。一天，他们决定散尽家财去边塞的沙漠里寻找一群在逃的犯人。那群人曾在木匠们的村子里发现一个古墓，并且盗走了里面的宝藏——一顶镶金丝的假发。木匠们想要夺回属于村民们的宝藏，还想要坏人们得到应有的惩罚。村民们对于他们这样的想法感到不可理喻：“这仨傻木匠，想啥呢！他们只是擅长做木工，怎么会抓坏人！”木匠们并没有在意，洒脱地踏上了翻山越岭的旅程，梦想着自己功成返乡那一刻的飒爽英姿，眼睛里闪着激动的光。那一霎，他们已不在意零散的送行人群，也不在意即将面对的风吹日晒和千辛万苦。塞下一口馒头，他们要做自己心里的英雄。

English story:

Once upon a time, in a southern village not far from here, lived a woman named Sarah Thacker who had got weak tharms. Together with her husband Simon and her big fat cat, she lived in a farm house with a thatched roof. Even when the windows were shut, there was always natural light shining throughout the house. The house wasn't much to look at inside, but it had a homely feel. There were shabby sofas, which when sat on, would sink right down to the floor, as they were so old. Simon was a shy but happy man, who was also a fabulous father. But on this specific day, he looked quite sad. This was because whilst he sat at his usual bench at the park, a group of fat thugs came up from behind him all of a sudden and pushed him over. He fell to the ground with a tremendous thud, and caught his thigh on a small sharp shard of shattered glass. The glass had been laying there from the previous weekend of fashionable party goers and drunk hooligans going too far with their drink. He felt like a shark had bitten him. A passer-by came to chase off the thugs and helped the man up. He thanked the man, and hobbled home where his wife bandaged his injury.

Appendix B Language Background Questionnaire

Participant Name *	**Experiment date & time**			
Gender	**Credit or payment?**		Credit	Payment
Age				
Place of Birth (City, Country)				
Native language/dialect	English	Mandarin		Cantonese
Is this also the first language you spoke after birth?	Yes	No		
Mother's native language/dialect				
Father's native language/dialect				
Which other languages/dialects do you speak?				
Language **Age of acquisition** **Competence:** Beginner/ Intermediate/Advanced/ Fluent		**Language** **Age of acquisition** **Competence:** Beginner/ Intermediate/Advanced/ Fluent		
Language **Age of acquisition** **Competence:** Beginner/ Intermediate/Advanced/ Fluent		**Language** **Age of acquisition** **Competence:** Beginner/ Intermediate/Advanced/ Fluent		
Please state the cities where you have lived for more than 6 months				
City, Country Years Months		**City, Country** Years		Months
City, Country Years Months		**City, Country** Years		Months
Are you right or left handed?		Right	Left	
Do you have (a history of) impaired hearing?		Yes	No	
If yes, please specify				
Do you have (a history of) learning and/or language impairment?		Yes	No	
If yes, please specify				

Do you have (a history of) a neurological disorder?	Yes	No
If yes, please specify		
Do you have impaired vision even when wearing glasses or contacts?	Yes	No
If yes, please specify		

*Following the General Data Protection Regulation 2018, participant name was no longer requested in the questionnaire since April 2018.