UNIVERSITY OF ALBERTA

The Production of English Fricatives by Native Mandarin Speakers

BY

Xiaozhen Zeng

A THESIS

SUBMITTED TO THE FACULTY OF ARTS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

BACHELOR OF ARTS

DEPARTMENT OF LINGUISTICS

EDMONTON, ALBERTA

April, 2023

UNIVERSITY OF ALBERTA

FACULTY OF ARTS

ABSTRACT

Learners of English may be influenced by the sounds of their mother tongue when speaking English. This study analyzes the acoustic features of English fricatives, such as the *s* sound in *hiss,* produced by native Mandarin speakers and native English speakers. This study also investigated how speaker groups differ and discussed the effects of individual variations on acquiring native-like pronunciation, and how their attitude toward second language acquisition factors may affect pronunciation.

Audio recordings were collected from 20 native Mandarin speakers and 10 native English speakers. Mandarin speakers were asked to fill out a brief questionnaire regarding their time living in English-speaking countries, their motivation for acquiring native-like pronunciation, the intensity of attention paid when speaking English, the extent of satisfaction with pronunciation, and the frequency of speaking English daily. Analysis revealed native Mandarin speakers can learn and produce English fricatives effectively and are more variable in place of articulation compared to their native English-speaking peers. Participants who rated themselves as highly motivated are more likely to be less satisfied with their pronunciation but produced fricatives less differently than native English speakers.

As the number of Mandarin-speaking international students continues to grow across Canadian secondary and post-secondary schools, the results from this study may inform second language teachers on how native Mandarin speakers articulate English fricatives.

ACKNOWLEDGEMENTS

I would like to thank my thesis supervisor, Dr. Benjamin V. Tucker for his support and guidance throughout the project, and Dr. Marina Blekher for the feedback as the second reader. I would also like to thank the members of the Alberta Phonetics Lab. I would like to express my gratitude to my participants for their patience. I am grateful to my friends and family for their emotional support. I am grateful to Susumu Hirasawa, I could not have written the thesis without listening to his music. I am also grateful to my cover music pals for making fantastic cover songs.

TABLE OF CONTENTS

TABLES AND FIGURES

CHAPTER 1. ENGLISH FRICATIVES AND SECOND LANGUAGE ACQUISITION

1.0 Spectral Properties of English Fricatives

Fricatives are generally known as the hissing sound produced by creating a narrow constriction in the oral cavity, forming turbulence of air that results in frication noise (Jongman et al., 2000). There are nine fricatives in the English sound inventory, and they are usually categorized by their place of articulation, paired up in voicing: labiodental /f,v/, interdental /θ,ð/, alveolar /s,z/, postalveolar /ʃ,ʒ/, and glottal /h/ (with no voiced counterpart).

Extensive research has been performed on the spectral properties of fricative sounds and which one(s) of them could effectively differentiate the sounds from each other. Early in 1988, Forrest et al. derived four spectral moments for studying a small corpus of syllable-initial fricatives. Spectral moment analysis based on a fast Fourier transformed (FFT) spectrum involves using statistical features to identify fricative sounds. Spectral mean, later known as the center of gravity, together with spectral variance (standard deviation), capture the local mean frequency and the distribution. Spectral skewness and kurtosis carry information on global tiltedness and peakedness. Spectral tilting normally reveals the energy distribution of the examined sound. More specifically, a positive skewness value indicates a negative tilt, where the energy concentrates in the lower frequencies, and vice versa. Kurtosis has been associated with the distribution of peaks; a high kurtosis value indicates more peaks in the FFT window and a more clearly defined spectrum with well-resolved peaks. The classification rate was good for sibilants but not so desirable for non-sibilants. However, this legacy of spectral moments inspired further studies on fricative classification. For instance, Shadle and Mair (1996) recorded two subjects whose native languages were American English and French respectively,

presented with tokens in vowel-fricative [VF] and stop-vowel-fricative-vowel [pV1FV2] styles. Spectral moments and spectral slope were taken from the onset, midpoint, and offset of the fricative segments, and reported that spectral mean, skewness, and kurtosis changed more across fricatives. The study was limited by the sampling frequency range, as the sampling frequency topped at 17kHz.

A more recent study that thoroughly examined spectral moments was done by Jongman et al. (2000), which had slightly different results than Shadle and Mair. This study recorded twenty native English speakers reading words containing labiodental, interdental, alveolar, and postalveolar fricatives. All spectral moments were able to distinguish places of articulation to different extents. Alveolar fricatives /s,z/ had the highest spectral mean frequencies at around 6133 Hz, postalveolar fricatives /ʃ,ʒ/ had the lowest at around 5108 Hz, and the dental fricatives laid in between with no significance. Spectral variance was low for sibilants and high for non-sibilants, again it failed to show the difference between the non-sibilants. Analysis on spectral skewness revealed that it was able to differentiate all places of articulation. Kurtosis failed to distinguish /f,v/ from /s,z/, but significance was obtained for other comparisons. And the other spectral feature, spectral peak frequencies were able to distinguish all places of articulation.

The other dimension that will also be considered in this study of fricatives is voicing. Also showed in Jongman et al. (2000), voiceless segments showed a higher spectral mean, lower spectral variance, higher skewness and higher kurtosis than their voiced counterparts. Kharlamov et al. (2022) showed that segment duration was longer for the voiceless fricatives in their study on temporal and spectral characteristics of fricatives in careful and conversational speech, this result agreed with many other previous findings. Furthermore, the results also showed that segment duration was more robust in showing the difference in careful speech than

in conversational speech, which indicated that duration is a major cue for voicing in careful speech.

1.1 Mapping Mandarin and English Fricatives

Standard Chinese, or Putonghua, is spoken widely across Mainland China regardless of whether one was born speaking a different dialect of Chinese. There are ten sounds that can be classified as fricatives and affricates in the Mandarin Chinese sound inventory as demonstrated in Table 1.

Table 1. Pinyin and corresponding IPA transcription and place of articulation (based on San, 2007)

Place of articulation	*Pinyin*	IPA transcription
Labiodental	f	/f/
Dental	z	/ts/
	c	/tsʰ/
	s	/s/
Retroflex	zh	/ʈʂ/
	ch	/ʈʂʰ/
	sh	/ʂ/
	r	/ʐ/
Alveolo-palatal	x	/ɕ/
	j	/tɕ/
	q	/tɕʰ/
Velar	h	/x/

Various spectral properties are used to measure Mandarin fricatives and affricates. Li S. and Gu (2015) studied alveolar, alveolo-palatal, and retroflex affricates /ts, tsʰ, tɕ, tɕʰ, ʈʂ, ʈʂʰ/ using spectral measures similar to Jongman et al. (2000). They found consistent effects on the spectral peaks, spectral mean (m1), and skewness (M3) that the values decrease in the order of alveolar, alveolo-palatal, retroflex. This was due to the back constriction that lengthened the front oral cavity, resulting in a lower frequency.

Because English and Mandarin have such different sound inventories, it is likely

inevitable for native Mandarin speakers who learn English as their second language to transfer the sounds in their mother tongue to replace English sounds that they are not familiar with. The cross-linguistic mapping of sounds in English and Mandarin, as shown in Table 2, represents the possible transfer of sounds connecting with lines. Besides /f/ and /s/ that show up in both languages and are phonemically identical, other fricative sounds are indirectly mapped to neighbours that shift in place or manner of articulation. The postalveolar fricatives /ʃ,ʒ/ are mapped to retroflex fricatives /ʂ,ʐ/, shifting in place of articulation (Liu S., 1990). /z/ is mapped to /ts/, since *pinyin* transcribes /ts/ as the letter *z*. For the English fricatives which lack correspondence in Mandarin, interdental /θ,ð/ were more likely to be replaced by /s,ts/ (Liu, N., 1988, Liu, S.,1990, Ma, 2019). The transfer phenomenon is also observed on the labiodental fricative /v/, which is often replaced by the labio-velar approximant /w/. All these differences in sounds make it harder for learners to master English sounds.

Table 2. The cross-linguistic mapping of English and Mandarin fricatives with Pinyin transcription

	English fricatives	Mandarin sounds	*Pinyin*
labiodental	f v	f	*f*
interdental	θ ð		
alveolar	s z	s	*s*
postalveolar	ʃ ʒ		
retroflex		ʂ ʐ	*sh* *zh*
velar		x	*h*
glottal	h		
affricate		ts	*z*
approximant		w	*w*

1.2 Second Language Acquisition

Behaviourism and nativism are two opposing theories of learning dominated the debate in early second language acquisition (SLA) theories. Behaviourists believed that learning a language results from a conditioned response, linked with rewarding and punishing for an action (VanPatten et al., 2020), where the response from the environment plays a crucial role in the process. Nativists held different views that language acquisition is the consequence of exposure to and immersion in a new language (Hummel, 2021).

In the naturalistic SLA, the age factor and other individual differences also play a role in success. The critical period hypothesis (CPH) proposed by Lenneberg in 1967 suggested a "built-in biological schedule" for language acquisition that one's ability to acquire language decreases after puberty. Though other research by Krashen (1973), Scovel (1984) and more recent studies suggested different cut-off points, a biological critical period exists for language acquisition. Pronunciation as a part of SLA is also shown to be closely linked with age (Flege et al., 1999). The research pointed out that new phonetic contrasts are processed through a first language (L1) filter, whereby having learned to pronounce the L1 too well results in having an accent. However, a later study showed that it is not impossible for highly motivated post-critical-period learners to achieve a native-like pronunciation immersed in the L2 environment with training in perceiving and producing speech sounds (Bongaerts et al., 2000).

1.3 Research Objectives

The first part of this project is focused on collecting a corpus of careful speech in English by native Mandarin speakers and native English speakers, consisting of English fricatives presented in syllables and natural English words. This will enable further research on phonetic traits of L1 Mandarin L2 English speakers.

There are two main objectives for the second part of this project, which concerns the statistical analysis of spectral and temporal features of speech produced by native Mandarin speakers in comparison to their native English-speaking peers:

1. To find out if there are differences in the production of English fricatives between the two speaker groups.
2. To look for the correlation between fricative production and the factors influencing second language acquisition (i.e., which factor(s) could act better at predicting more native-like speech sound).

CHAPTER 2. DATA COLLECTION

2.0 Participants

Thirty speakers are recruited from the University of Alberta Linguistics subject pool. Twenty of them are native Mandarin speakers who were learners of English, and ten are monolingual English speakers who grew up in western Canada, speaking only English at home. Participants were granted course credit for their participation.

2.1 Materials

Nine English fricatives /f,v,θ,ð,s,z,ʃ,ʒ,h/ were recorded in vowel-consonant-vowel (VCV) syllables and real English words. The fricatives were in the medial position for the syllables, flanked by a pair of identical vowels from /i,a,u/. Three English words were selected for each word-initial, word-medial, and word-final position. Exceptions applied to /ʒ/ and /h/, as American English tends to produce word-initial and word-final /ʒ/ as its correlated affricate /dʒ/, and no words in English contain word-final /h/. Each token was repeated three times. Altogether this yielded a total of 297 tokens attached in Appendix C (9 fricatives × 3 vowels × 3 repetitions, 9 fricatives × 3 words × 3 positions × 3 repetitions – exceptions).

A questionnaire (attached in Appendix D) consisting of three types of questions (adapted from Liao 2006) was filled out by native Mandarin speakers. The first type of questions asked about basic demographic information such as the year of study and time spent in English-speaking countries. The second type of question asked about the participant's attitude toward their English accent and pronunciation.

2.2 Procedure

After a brief introduction and signing a consent form, participants who identified themselves as native Mandarin speaker were asked to fill out the questionnaire. The participants then were

recorded in a sound-attenuated whisper booth in the Alberta Phonetics Laboratory using the KORG MR-2000s studio recorder. A head-mounted microphone (Countryman E6) was attached to the participant approximately 3 cm away from the left corner of the participant's mouth throughout the experiment. The instructions along with a list of syllables and words were presented through PowerPoint slides on a computer which can be controlled by the participant.

For the syllables, a pronunciation guide indicating the sounds using English words as well as the corresponding IPA transcript were provided on the same slide. Prior to the actual recording, the researcher ran through the manner of pronouncing syllables and made sure that the participants understood the instructions properly. A few examples were given to the participants for practice purposes until they felt comfortable.

2.3 Measures

Recordings were manually annotated with fricative segments and words in Praat (Boersma and Weenink, 2020), and using a custom script to extract the acoustic parameters from annotated segments. Prior to extracting acoustic measures, all audio recordings were down-sampled to 16000 Hz. A set of spectral and temporal parameters was chosen among many possible cues for fricatives (Jongman et al., 2000; McMurray and Jongman, 2011; Kharlamov et al., 2022). Four spectral moments, namely spectral mean (also known as "centre of gravity", spectral variance (also known as "standard deviation"), skewness, and kurtosis were measured and can serve as indications for place of articulation and voicing. Spectral moments were measured with a 40 ms Hamming window at three different locations (onset, midpoint, and offset). Spectral peak, the tallest peak in the fast Fourier transform (FFT) spectrum, was also measured in Hertz using a 40 ms Hamming window. Duration as a temporal measure for indicating voicing was measured.

2.4 Data Analysis

Data analysis was performed in R 4.2.1 (R Core Team, 2022). First, for acoustic parameters, linear mixed-effects models were conducted with the acoustic measures as the dependent variables. This was done using the LME4 package 1.1-30 (Bates et al., 2015) and the lmerTest package (Kuznetsova et al., 2017). Each model examines the fixed effects of language group (English versus Mandarin) and place of articulation (labiodental, interdental, alveolar, postalveolar, glottal). The intercepts for the factors in the models were set as follows: "English" for language group, "labiodental" for place of articulation. To better understand the similarities and differences between the production of two language groups, each segment was examined with acoustic measures as a function of language groups, with "English" set as intercept.

The other goal of this project was to find out how native Mandarin speakers performance on pronunciation compared to that of native English speaks, and how is the performance correlate with self-rated second language acquisition factors. First, the inter-relationship between the self-rated SLA factors was examined using fitting linear model in R stats package. Linear models were fitted with one of the factors as dependent variable, back fit to find the best fit assisted by Akaike information criterion (AIC). This process was repeated for all collected factors. Next, for linking up the acoustic parameters and SLA factors, linear mixed-effects models were conducted with the factors as predictors and acoustic measures as dependent variables. This was done again using the LME4 package 1.1-30 (Bates et al., 2015) and the lmerTest package (Kuznetsova et al., 2017). Each model examines the fixed effects of language group (English versus Mandarin), and levels of self-rating within Mandarin speaking group. The intercept was set as 'English group' for the factors in the models.

CHAPTER 3. RESULTS

3.0 Spectral and temporal properties

3.0.1 Spectral mean

For ONSET SPECTRAL MEAN, as shown in Figure 1(a), the alveolar fricatives had the highest frequency among other sounds at around 5000 Hz for Mandarin speakers and 4700 Hz for English speakers, and the labiodental was the lowest at around 3600 Hz and 3600 Hz for Mandarin and English speakers respectively. Interdentals (at around 4000 Hz for both groups) and postalveolars (3800 Hz for the Mandarin group and 4000 for the English group) fell in between. The mean for /h/ laid at around 3800 Hz for both groups. As for the language group comparison, no overt effect was found. Yet a post hoc test investigating the effect of language for individual segments showed that the onset of /v/ was different between the two language groups, with Mandarin speakers having lower ONSET SPECTRAL MEAN than English speakers (β=-485.64, t=-2.55, p<0.05). A minor difference was observed for /ʒ/, with Mandarin speakers at a higher frequency than English speakers (β=-305.265, t=-1.955, p=0.061).

For MIDPOINT SPECTRAL MEAN, the alveolar fricatives were the highest at 5500 Hz for Mandarin speakers and 5800 Hz for English speakers. However, unlike what was seen for ONSET SPECTRAL MEAN, the labiodentals were the lowest for Mandarin speakers at around 4200 Hz (4500 Hz for English speakers), and postalveolars were the lowest for English speakers at 4500 Hz (4400 Hz for Mandarin speakers). Interdentals fell in between at 4900 Hz for Mandarin speakers and 4600 Hz for English speakers. The glottal /h/ was at 3500 Hz for both groups. A significant effect was observed in comparing two language groups, where the frequency of midpoint spectral mean for Mandarin speakers was lower than English speakers (β=-232.50, t=-2.30, p<0.05). Post hoc tests further showed that /v/ produced by Mandarin speakers was lower

than English speakers (β=-467.49, t=-2.41, p<0.05). A minor effect also showed for /z/, with Mandarin speakers being marginally lower than English speakers (β=-402.30, t=-1.97, p=0.59).

For OFFSET SPECTRAL MEAN, again, alveolars were at the highest frequency at around 4500 Hz for both groups, and labiodentals were at the lowest at around 3300 Hz for both groups. Interdentals and postalveolars fell in between. Interdentals (3900 Hz) were higher than postalveolars (3700 Hz) for Mandarin speakers, while English speakers had the other way around, with postalveolars (3800 Hz) being higher than the interdentals (3250 Hz). No overt difference was shown comparing the two groups. A main effect was seen for the interdental /θ/, with Mandarin speakers at higher frequencies (β=467.21, t=3.52, p<0.01). Minor effects were seen for /f/, with Mandarin speakers at a higher frequency than English speakers (β=236.22, t=1.88, p=0.07), as well as for /ð/ (β=380.16, t=2.017, p=0.05).

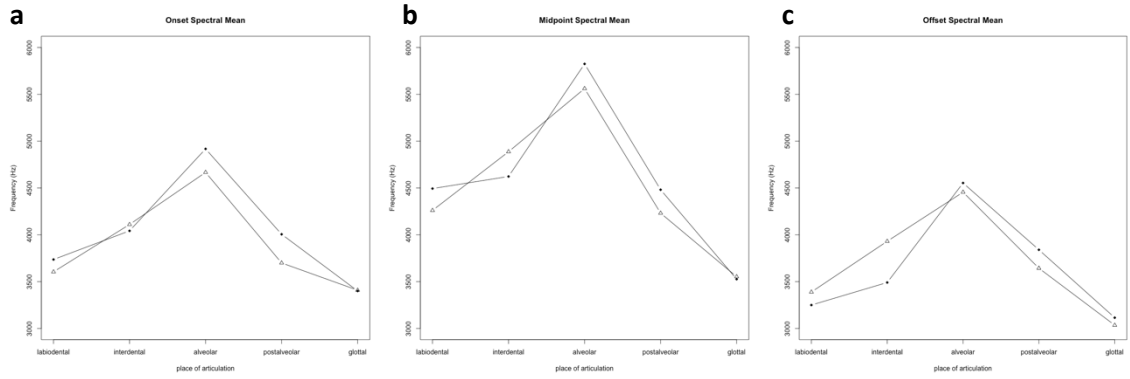


Figure 1. SPECTRAL MEAN at onset (a), midpoint (b), and offset (c) positions as a function of place of articulation. Native Mandarin-speaking group is illustrated with triangles and diamonds as native English-speaking group.

3.0.2 Spectral variance

For ONSET SPECTRAL VARIANCE, as shown in Figure 2(b), the Mandarin group had the largest range for labiodentals at 2200 Hz, and the lowest for postalveolar at 1900 Hz, with interdentals (2150 Hz) and alveolars (2000 Hz) falling in between. English speakers had the

most variable range for interdentals (2250 Hz) and the least variable for postalveolar (1800 Hz), with labiodentals (2200 Hz) and alveolars (1950 Hz) falling in between. The glottal /h/was at 2000 Hz for both groups. No overt significance was found between the two groups. However, post hoc tests further revealed that there were main effects seen for /f/ (β=49.50, t=2.10, p<0.05), /z/ (β=215.93, t=4.48, p<0.001), /ʃ/ (β=86.52, t=2.537, p<0.05), and /ʒ/ (β=134.01, t=3.33, p<0.01), with Mandarin speakers producing these sounds more variably than English speakers. Mandarin speakers produced /ð/ (β=-144.65, t=-3.15, p<0.01) less variably than English speakers. A minor effect was seen for /s/, with Mandarin speakers being marginally more variable (β=-81.01, t=2.035, p=0.05).

For MIDPOINT SPECTRAL VARIANCE, as shown in Figure 2(b), Mandarin speakers showed the most variable production for the labiodentals (2000 Hz), and the least variable production for the alveolars (1720 Hz), with interdentals (2000 Hz) and postalveolars (1700 Hz) falling in between. For English speakers, labiodentals and interdentals showed similar variability (2100 Hz), followed by postalveolar at 1600 Hz. Alveolar showed the least variability at 1460 Hz. The glottal /h/ was at 1850 Hz for Mandarin speakers, and 1800 Hz for English speakers. A main effect was obtained comparing between two groups, with Mandarin speakers produced fricative more variably than English speakers (β=71.34, t=2.50, p<0.05). Major effects were seen in /ð/ (β=-152.54, t=-2.71, p<0.05), with Mandarin speakers being less variable than English speakers. Effects also obtained for /s/ (β=181.90, t=2.87, p<0.01), /z/ (β=352.63, t=5.42, p<0.001), /ʃ/ (β=166.35, t=3.34, p<0.01), /ʒ/ (β=184.95, t=4.55, p<0.001), with Mandarin speakers being more variable. Minor effects were shown for /θ/ (β=-105.92, t=-1.94, p=0.06), where Mandarin speakers were less variable in producing the sound.

For OFFSET SPECTRAL VARIANCE, as illustrated in Figure 2(c), Mandarin speakers showed

the most variability in producing the interdentals (2170 Hz), and the least variable in producing the postalveolars (1930 Hz). The labiodentals and alveolars laid at around 2140 Hz. English speakers showed the most variable in producing interdentals at 2030 Hz, and the least in postalveolars (1800 Hz), with labiodentals (2150 Hz) and alveolars (2030 Hz) falling in between. The glottal /h/ was at 1920 Hz for Mandarin speakers, and 1940 Hz for English speakers. No overt effect was observed between the two groups. Further, in post hoc tests, main effects were seen in /θ/ (β=-84.89, t=-2.33, p<0.05), where Mandarin speakers produced the sound less variable than English speakers. Effects for /z/ (β=146.45, t=3.72, p<0.001), /ʃ/ (β=125.98, t=3.37, p<0.01), /ʒ/ (β=113.86, t=3.01, p<0.01), revealing that Mandarin speakers were more variable in producing these sounds.

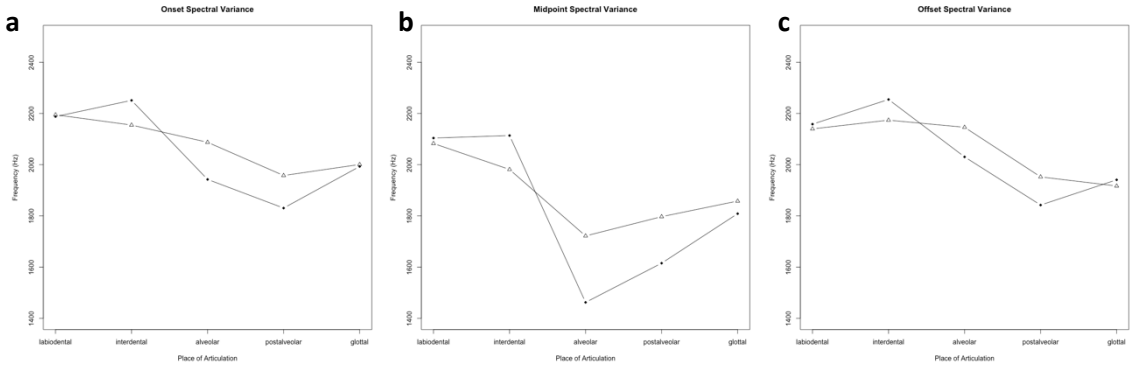


Figure 2. SPECTRAL VARIANCE at onset (a), midpoint (b), and offset (c) positions by language group, triangle as Mandarin speakers and diamond as English speakers.

3.0.3 Spectral skewness

For ONSET SPECTRAL SKEWNESS, as illustrated in Figure 3(a), the interdentals, labiodentals, and alveolars had increasingly negative skewness values for both groups, indicating energy concentrated in the higher frequencies. And postalveolars, as well as the glottal /h/, had positive skewness values, where the energy concentrated in the lower frequencies. No overt effect was observed between the two groups. In post hoc tests, /z/ (β=0.38, t=3.08, p<0.01), /ʒ/ (β=0.20,

t=2.20, *p*<0.05), with Mandarin speakers had skewness values more positive than English speakers, where these sounds were produced lower in frequencies.

For MIDPOINT SPECTRAL SKEWNESS, as shown in Figure 3(b), it had the same trend but spanned a larger value range than the onset skewness. No overt effect was seen between the two groups. Post hoc tests revealed that the energy of /v/ produced by Mandarin speakers had energy concentrated in lower frequencies than English speakers (*β*=0.28, *t*=2.33, *p*<0.05).

For offset spectral skewness, as shown in Figure 3(c), again the trend was the same as mentioned for skewness at the other two positions. No overt effect was seen between the two groups. Further, in post hoc tests, main effects were seen in /f/ (*β*=-0.18, *t*=-2.24, *p*<0.05) and /θ/ (*β*=-0.27, *t*=-3.36, *p*<0.01), indicating these two sounds produced by Mandarin speakers had lower skewness values than English speakers, thus the energy concentrated in higher frequencies.

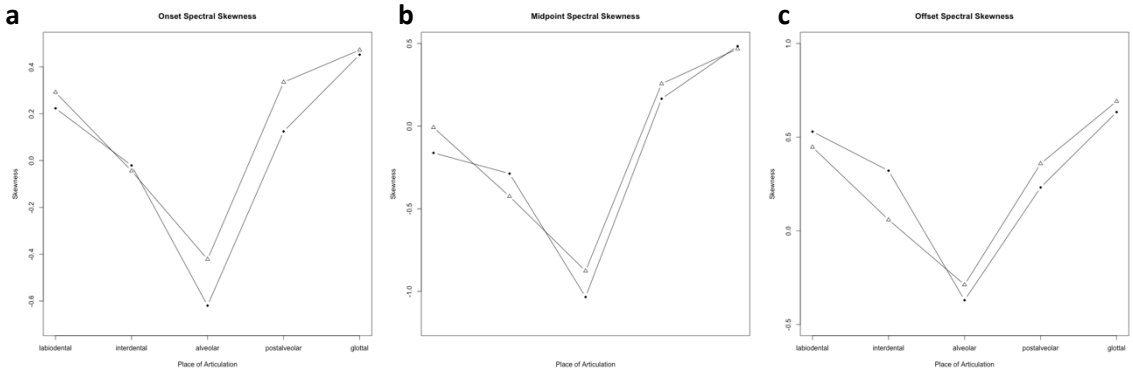


Figure 3. SPECTRAL SKEWNESS at onset (a), midpoint (b), and offset (c) positions by language group, triangle as Mandarin speakers and diamond as English speakers.

3.0.4 Kurtosis

For ONSET KURTOSIS, as shown in Figure 4(a), the alveolars had the highest kurtosis value among other sounds produced by both language groups, with -0.40 for Mandarin speakers and 0.14 for English speakers. Mandarin speakers had the lowest kurtosis value shown for

interdentals at -0.70, labiodentals (-0.69), postalveolar (-0.42) fell in between. While for English speakers, the lowest kurtosis value was observed at -0.92 for interdentals, labiodentals (-0.74) and postalveolars (-0.42) fell in between. The glottal /h/ was -0.44 for Mandarin speakers and -0.32 for English speakers. No overt effect was obtained between the two language groups. However, /ð/ (β=0.37, t=3.29, p<0.01) showed more peaks in spectra extracted from Mandarin speakers than that from English speakers. On the opposite, spectra extracted from Mandarin speakers for /z/ (β=-0.87, t=-4.80, p<0.001) were less peaked than that from English speakers.

For MIDPOINT KURTOSIS, as shown in Figure 4(b), both language groups had the highest values for the alveolars (0.67 for Mandarin speakers and 1.76 for English speakers) and the lowest for the interdentals (-0.31 and -0.12, respectively). The other sounds fell in between. A minor effect was seen comparing the two language groups, with Mandarin speakers having lower kurtosis values than English speakers (β=-0.17, t=-1.98, p=0.06). Post hoc test further revealed that main effects were found in /ð/ (β=0.51, t=3.11, p<0.01), indicating more peaked spectra were extracted from Mandarin speakers than that from English speakers. /s/ (β=-0.79, t=-2.64, p<0.05), /z/ (β=-1.41, t=-4.56, p<0.001), /ʃ/ (β=-0.23, t=-3.77, p<0.001), and /h/ (β=-0.27, t=-2.70, p<0.05) on the other hand, showed less peak on spectra extract from Mandarin speakers.

For OFFSET KURTOSIS, as shown in Figure 4(c), /h/ was the highest for both groups, with 0.04 for Mandarin speakers and 0.15 for English speakers, and interdentals being the lowest (-0.70 and -0.77 respectively). The values for other sounds fell in between. No overt effect was shown comparing the two groups. However, the spectra of /f/ (β=-0.25, t=-2.22, p<0.05) and /z/ (β=-0.29, t=-2.34, p<0.05) extracted from Mandarin speakers were shown to be more flattened out than that from English speakers.

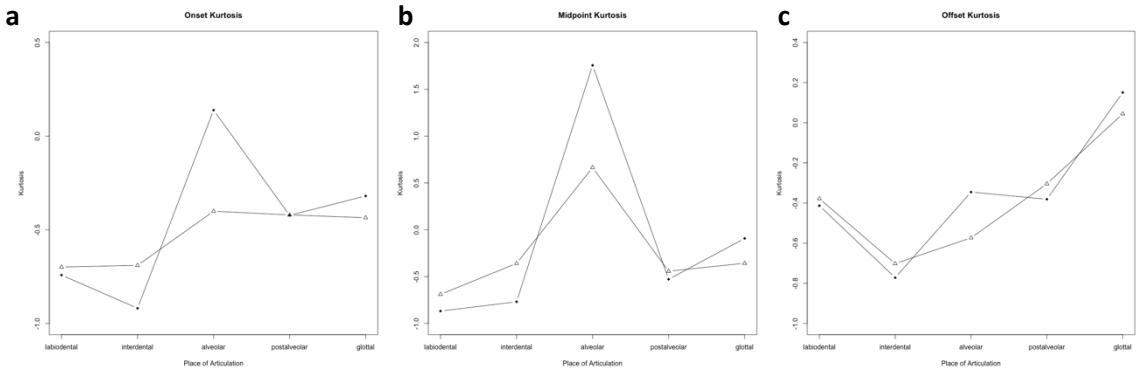


Figure 4. SPECTRAL KURTOSIS at onset (a), midpoint (b), and offset (c) positions by language group, triangle as Mandarin speakers and diamond as English speakers.

3.0.5 Spectral peak

A main effect was found for spectral peaks between the two speaker groups, where Mandarin speakers had lower spectral peaks in general than English speakers. The overall tendency of spectral peak for Mandarin speakers was not consistent with Jongman's finding, alveolars had the highest peak, and glottal was the lowest, as shown in Figure 5(a). Interdentals, labiodental, and postalveolars lay in between. English speakers showed the same trend. No overt effect was seen between the two groups.

/v/ showed significance in Post hoc tests, the peak frequency derived from Mandarin speakers was much lower than from English speakers (β=-1090.21, t=-2.57, p<0.05). Even so, as indicated in previous studies, the dental-related fricatives were characterized by relatively flat spectra, thus no clear dominating peak had been found in any frequencies. The significance of /v/ may suggest the difference in place of articulation between the two groups.

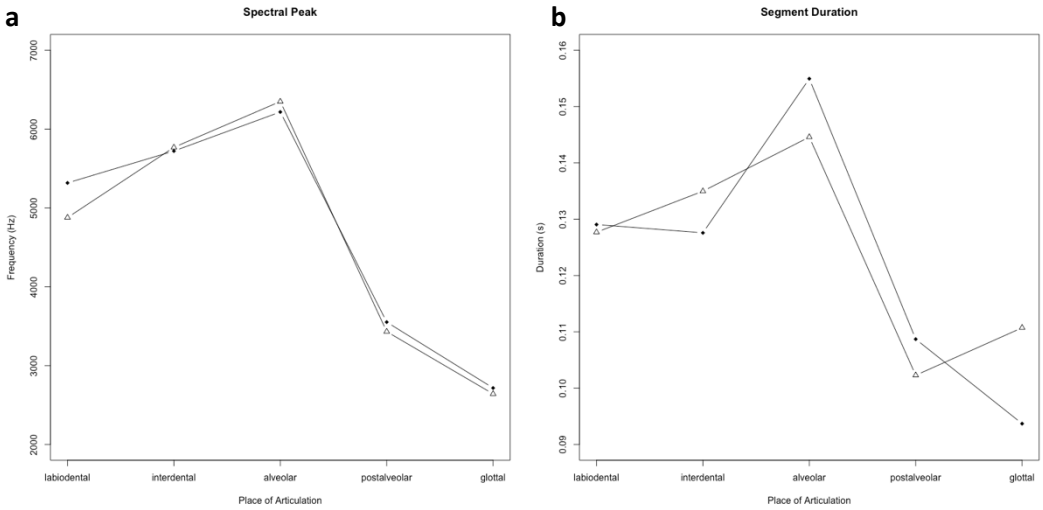


Figure 5. SPECTRAL PEAK (a) and SEGMENT DURATION (b) by language groups. Native Mandarin-speaking group is shown in triangle and native English-speaking group is shown in diamond.

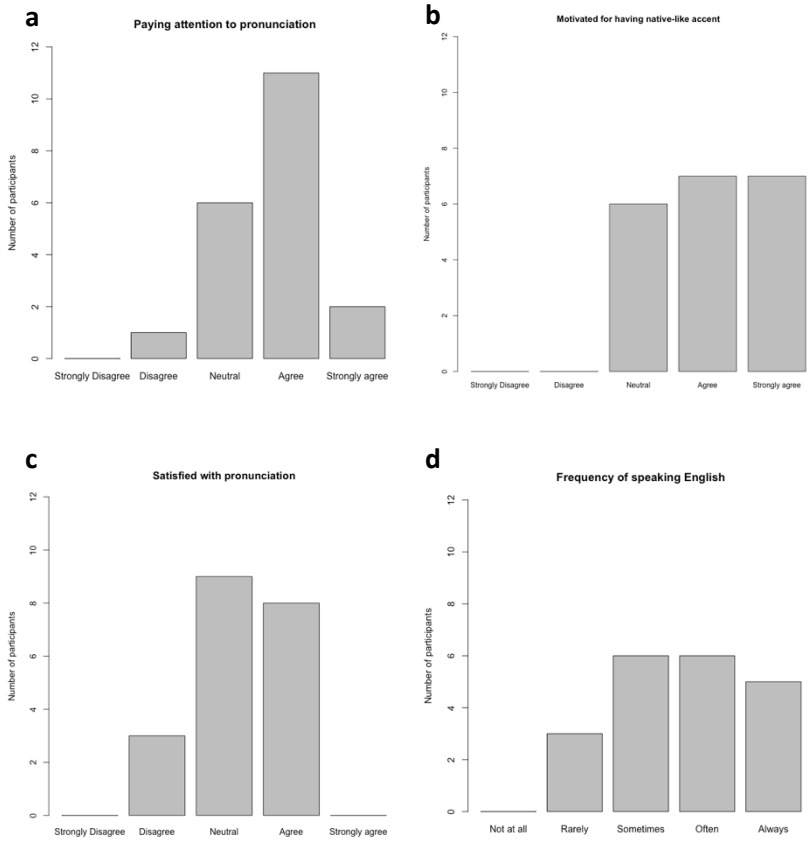
3.0.6 Duration

No main effect was shown for duration between two groups. Post hoc test showed that Mandarin speakers produced /z/ significantly shorter than English speakers (β=-0.02, t=-2.27, p<0.05). Duration could reliably differentiate voicing, with voiceless sounds being longer than the voiced sounds (β=0.04, t=34.55, p<0.001). To investigate the hypothesis that Mandarin speakers produced less voiced /z/, a test was conducted for language group and voicing at the place of articulation of alveolar, with intercept set as English and /s/. The results showed that /z/ produced by Mandarin speakers was marginally different or no difference from /s/ produced by English speakers (β=-0.009, t=-1.74, p=0.08).

3.1 Second language acquisition factors

The distribution of self-ratings is shown in Figure 6. The rating ranged from disagree to strongly agree for "paying attention to your pronunciation" (abbreviated as *attention* later, Figure 6a), "motivated for acquiring native-like accent" (abbreviated as *motivation*, Figure 6b)

and “satisfied with your pronunciation” (abbreviated as *satisfaction*, Figure 6c), with more participants who would like to rate themselves from neutral to strongly agree than the negative extreme in general. Cumulatively, over half of the participants graded their *frequency of speaking English* (abbreviated as FreqSpeak, Figure 6d) as sometimes and often, 3 rated rarely, and 5 rated always. For *self-rated proficiency* (Figure 6e), the rating ranged from 2 to 5, half of the participants rated 3 on their proficiency. The time the participants had stayed in Canada (abbreviated as *LiveCa* ,Figure 6f) varied from a few months to eight years and was distributed almost evenly.



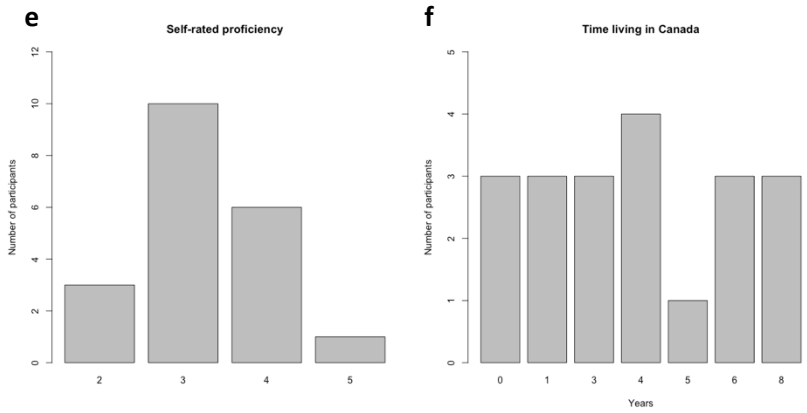


Figure 6. The distribution of self-rating of second language acquisition factors.

Simple linear regression models were run to investigate the correlation of individual factors with each other. The model revealed that *satisfaction* with pronunciation is positively correlated to the intensity of *attention* paid to pronunciation when speaking English (R^2=0.27, F(1, 18)=7.87, p<0.05). *Self-rated proficiency* showed a significant positive correlation with the *frequency* of speaking English (R^2=0.28, F(1, 18)=8.357, p<0.01). Other factors had no effect on each other (all ps >0.1).

Multiple linear regression models were calculated and backward fitted to find the best fit predicting each individual factor by the other factors. A significant effect was found in the regression equation predicting *satisfaction* based on *attention*, *motivation*, and *frequency* (R^2=0.45, F(3, 16)=6.13, p<0.01). *Attention* (β=0.67, t=3.84, p=0.001) and *motivation* (β=-0.33, t=-2.16, p<0.05) had the main effects on predicting *satisfaction*, with *frequency* (β=0.21, t=1.76, p<0.1) having a minor effect. For predicting *attention*, significance was found in the regression equation based on *motivation* and *satisfaction* (R^2=0.41, F(2, 17)=7.502, p<0.01), with *motivation* (β=0.37, t=2.30, p<0.05) and *satisfaction* (β=0.65, t=3.52, p<0.01) both contributing significantly to the model. For predicting *time living in Canada* (*LiveCa*), *frequency*, and *self-rated proficiency*, no best-fit multiple linear regression model was found.

Based on results from the last part and previous research, six acoustic parameters, spectral moments (SPECTRAL MEAN, SPECTRAL VARIANCE, SKEWNESS, KURTOSIS) at MIDPOINT position, SPECTRAL PEAK, and DURATION were selected for examining the correlation with second language acquisition factors.

For *attention*, effects were seen in MIDPOINT SPECTRAL VARIANCE (β=86.90, t=2.40, p<0.05), and SPECTRAL PEAK (β=-470.50, t=-2.17, p<0.05) for the neutral category. Minor effects were obtained in MIDPOINT SPECTRAL VARIANCE for agree category (β=59.39, t=1.94, p=0.06), and MIDPOINT KURTOSIS for neural (β=-0.23, t=-2.02, p=0.05). The neutral category showed more difference than the agree category, and no difference was obtained in disagree and strongly agree categories, indicating that participants who rated themselves as neutral on *attention* produced fricatives more variably than English speakers.

For *motivation*, effects were seen in the MIDPOINT SPECTRAL VARIANCE for the neutral (β=84.89, t=2.42, p<0.05) and agree (β=93.75, t=2.79, p<0.01) categories. The agree category also showed a minor effect in the SPECTRAL MEAN (β=-230.78, t=-1.93, p=0.065). As main effects were obtained mainly in the midpoint spectral variance, the agree category had a higher significance level than the neutral category, suggesting that participants who rated agree were more variable in producing English fricatives.

For *satisfaction*, main effects were obtained in MIDPOINT SPECTRAL VARIANCE for agree (β=90.70, t=2.75, p<0.05) and DURATION for agree (β=0.02, t=2.09, p<0.05). Minor effects were seen in SPECTRAL VARIANCE for neutral (β=64.84, t=2.03, p=0.05), KURTOSIS for neural (β=-0.20, t=-1.94, p=0.06) and agree (β=-0.21, t=-1.96, p=0.06), and SPECTRAL SPEAK for neutral (β=-375.29, t=-2.00, p=0.06). This indicated that the more satisfied the participants were on the pronunciation, the more different they were from native English speakers.

For *self-rated proficiency*, main effects were seen in SPECTRAL VARIANCE for rate 4 (β=92.52, t=2.60, p<0.05) and SPECTRAL PEAK for rate 3 (β=375.50, t=-2.14, p<0.05). A minor effect was found in MIDPOINT KURTOSIS for rate 5 (β=-0.46, t=-1.97, p=0.06). Participants who rated 3 or 4 showed more variability in place of articulation when producing English fricatives.

For *FreqSpeak*, main effects were seen in MIDPOINT SPECTRAL VARIANCE for sometimes (β=80,84, t=2.13, p<0.05), strongly agree (β=102.11, t=2.54, p<0.05), and MIDPOINT KURTOSIS for strongly agree (β=-0.29, t=-2.32, p<0.05). This suggested that participants who rated strongly agree produced fricatives more variably than those who rated neutral, and the spectra extracted from them showed different peakedness than native English speakers.

Finally, for *LiveCa*, main effects were found in spectral variance for residing length of 4 years (β=116.13, t=2.89, p<0.05), kurtosis for 6 years (β=-0.36, t=-2.41, p<0.05), and duration for 5 years (β=-0.04, t=-2.48, p<0.05). Minor effects were obtained in spectral variance for 5 years (β=152,85, t=2.01, p=0.06) and 6 years (β=96.71, t=2.01, p=0.06), as well as the duration for 4 years (β=0.02, t=2.014, p=0.06). This revealed that there was no overt tendency in how the length of living in Canada relates to the performance in pronunciation.

CHAPTER 4. DISCUSSION & CONCLUSIONS

4.0 Spectral Properties of Fricatives Produced by Two Speaker Groups

This study aimed to compare the production of English fricatives between native Mandarin speakers and native English speakers in terms of acoustic features. Spectral and temporal features were selected to identify the place of articulation and voicing of the speech sounds.

Spectral moments analyze speech sound from a statistical point of view, summarizing the sound in local (mean frequency) to global (tiltedness and peakedness) (Jongman et al., 2000). The main effect found in MIDPOINT SPECTRAL VARIANCE suggested that Mandarin speakers were more variable in producing English fricatives than their English-speaking peers. The production of /ð/ was less variable in the distribution of the mean frequency range, while /s,z,ʃ,ʒ/ were highly variable compared to native English speakers. A marginal effect was observed in KURTOSIS, and this indicated the more flattened spectra derived from Mandarin speakers than English speakers.

Temporal parameters, especially the SEGMENT DURATION, is known for its capability of indicating voicing of sounds. In this study, the voiceless and voiced segments are significantly different from each other in DURATION, with the voiceless being longer than voiced sound, which is expected as it is indicated in previous research. The only segment that showed significance between the two groups is /z/, with Mandarin speakers producing longer /z/. Further analysis confirmed that /z/ produced by Mandarin speakers could not be differentiated from /s/ by English speakers. It may be related to the fact that the letter z in Pinyin represents the voiceless affricate /ts/, and Mandarin speakers could easily borrow the sound when they were learning English. The production of alveolar affricate replacing voiced fricative is also confirmed by the research's notes in annotations.

A rating system could now be built based on how many parameters showed differences, graded out of a total of 14 parameters. /h/ had the lowest rating of difference (2/14). Although /h/ is more velarized than they are in English, it does not affect the identification of them being two separate sounds acoustically and statistically. /f,v,θ,s,ʒ/ are rated 3 out of 14. /f/ is different in the segment onset and offset, which may imply that the main body of this segment is not different between the two groups, but how native Mandarin speakers treat the transitioning from the preceding segment to the following segment is different from English speakers. Surprisingly, /θ, ʒ/ that do not have correlating or similar sounds in Mandarin are well performed, indicating that Mandarin speakers can effectively learn most sounds that are not in the inventory. /ð/ is rated 5 out of 14, this is reasonable because this sound is new for Mandarin speakers.

The voiced alveolar fricative /z/ is rated the highest with 9 out of 14. The significance in ONSET SPECTRAL MEAN may be related to the burst release of the stop /t/, as /z/ was hypothesized to be more likely being produced as /ts/ (Table 2). The effect seen in DURATION may be related to the voicing of this segment, which is indirectly confirmed by no significance obtained when comparing /z/ produced by Mandarin speakers to /s/ produced by English speakers. The discrepancy then arises as the minor effect in MIDPOINT SPECTRAL MEAN suggested that the frequency of /z/ produced by Mandarin speakers was lower than that of English speakers. This is contrary to the trending in the voicing of segments manifested in spectral mean, that voiceless segments have a higher frequency than the voiced segments. Nevertheless, /z/ produced by Mandarin speakers was significantly different from /z/ by English speakers, and it is remarkably different from the pattern of other fricatives varying between the two groups. This may be related to the way that other sounds in Mandarin could be indirectly mapped to phonemically neighbouring sounds while keeping the same voicing property. But for /z/, even if it has a

phonemic neighbour to be mapped to, differences in the manner of articulation (affricate to fricative) and voicing (voiceless to voiced) are still contributing to the significance observed between the two speaker groups.

The difference in producing /v/ related to indirect mapping was also mentioned in previous research on second language learning. English labiovelar approximant /w/ is often mispronounced as the labiodental approximant /ʋ/ or the labiodental fricative /v/, and is considered as a common mistake among Mandarin learners of English (Yue and Ling, 1994). Further in Ma's study (2019), she spotted that the English labiodental fricative /v/ was often mispronounced as the labiovelar approximant /w/ in students who speak Northern dialects and concluded as the negative transfer phenomenon in second language acquisition. In this study, the finding of /v/-/w/ replacement is indirectly confirmed with the spectral skewness in the Mandarin group being much more negative than the English group, indicating a concentration of energy in lower frequencies. This phenomenon was also spotted by the researcher when annotating the audio recordings.

4.1 Factors Influencing Second Language Acquisition

Unfortunately, due to the limitation on the sample size, the results for this section may not be reliably generalized to the greater public, the conclusions and discussion in this section are restricted to the participants in this study.

Results from this study are in line with previous research on factors influencing second language acquisition that high motivation is a crucial factor in language success (Richards, 1985). Participants who rated themselves as highly motivated produced more native-like fricatives and participants who had lower satisfaction with their pronunciation showed no

significant difference from native English speakers. The results for satisfaction are rather counterintuitive, that participants who were satisfied with their pronunciation showed more variability in place of articulation and voicing in the fricative production task than participants in other categories, and participants who were not satisfied with pronunciation did not show any significant difference in fricative production comparing to native English speakers. To examine the multiple regression model in which satisfaction and motivation are positively correlated with attention, it is successful in predicting the significance obtained for the neutral category but failed to predict the difference in the agree category. Since *satisfaction* and *attention* are positively correlated, satisfaction is able to predict that no significance was shown in the disagree category. The counterintuitive results that participants who did not pay much *attention* to pronunciation (disagree category) performed more native-like in the production of fricatives than those who rated neutral and agree could partially relate to Krashen's monitor hypothesis (Krashen, 1981). According to the original hypothesis, the monitor functions as a planner, editor, and corrector, when the learners have sufficient time to think about the correctness since they already knew the grammatical rules. Learners who use the monitor all the time are regarded as "monitor over-users", those who would prefer not to use the monitor are "monitor under-users", and "monitor optimal-users" are those who could utilize the monitor without over-complication. Although the original hypothesis was proposed for assessing grammar mastering, it is still useful in this study. Participants who were in the neutral and attentive categories could be monitor under-users, and they have not yet optimized their use of the monitor. Unsurprisingly, over-users performed better at producing English fricatives compared to monitor under-users. Most other participants rated as very attentive and inattentive may be categorized as optimal users, they show little difference in producing fricatives

compared to native English speakers.

Interestingly, participants who had a neutral self-rating (i.e., 3 in self-rated proficiency and neutral in other factors) were more likely to differ in the production of fricatives than native English speakers. It is possible that due to the ambiguity of the questionnaire design, the "neutral" option was understood as the getaway of having no reflection on the factors in learning and using English. The other possibility is that the participants did not have a clear stance on the matter, as the option originally intended to convey. Regardless of the possibilities, having a strong opinion on the factors is a positive indicator of acquiring a native-like accent.

It is worth noting that neither the time living in an English-speaking country nor the frequency of English speaking had an overt significance on pronunciation. Although individuals might hold various standards on sensing frequency, further exploration of the time immersing in an English environment suggested that the difference in pronunciation may be linked more closely to the participants as individuals, rather than the time and frequency of immersion. No tendency in time or frequency of listening to and speaking English reveals that the acquisition of a native-like accent may relate to other more subjective and affective factors such as *motivation* and *attention*. Again, due to the limitation on the size of the disagree (1 sample) and strongly agree (2 samples) categories, the results and conclusions discussed above may not be reliably generalized to the whole English-learning population.

4.2 Implications and Future Directions

In sum, the present study indicates that the production of English fricatives by native Mandarin speakers who learned English as a second language (ESL) was different from native English speakers in MIDPOINT SPECTRAL MEAN and MIDPOINT STANDARD DEVIATION in terms of

acoustic characteristics. This reveals that native Mandarin speakers produce English fricatives more variably in place of articulation than native English speakers. Breaking down to individual segments, the voiced alveolar /z/ was substantially different between the two groups. The differences include place of articulation, voicing, as well as energy and peak distribution on the spectrum, together with the researcher's notes during annotation, leading to the deduction of /z/-/ts/ replacement as mentioned in the cross-linguistic mapping (Table 2). The discrepancy of voicing manifested in MIDPOINT SPECTRAL MEAN should be further examined in later studies. Overall, future research will investigate how native English speakers and other learners of English perceive and understand the English fricatives produced by native Mandarin speakers.

Results of the present study in second language acquisition suggested that the time exposed to and immersed in the second language environment is not a crucial factor in acquiring a native-like accent. Rather, the attention paid to pronunciation when speaking English is more important. This may be a piece of information for ESL instructors, that for certain fricative sounds such as /z/, drawing students' attention to the place of articulation and manner of production can improve the likelihood of acquiring a native-like accent. Future research on this topic should increase the sample size to balance the number of ratings in each category, as well as examine more factors that may influence the attitude toward language learning.

REFERENCES

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. J.Stat.Soft, 67(1), 1-48.

Boersma, P., and Weenink, D. (2020). "Praat: Doing phonetics by computer (version 6.2.05) [computer program]," http://www.praat.org/ (Last viewed January 5, 2022).

Bongaerts, T., Mennen, S., & Slik, F. V. D. (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia linguistica*, *54*(2), 298-308.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of memory and language*, *41*(1), 78-104.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *The Journal of the Acoustical Society of America*, *84*(1), 115-123.

Hummel, K. M. (2021). Introducing second language acquisition: Perspectives and practices.

Jongman, A., Wayland, R., Wong, S. 2000. Acoustic characteristics of English fricatives. J. Acoust. Soc. Am. 108(3).

Kharlamov, V., Brenner, D., & Tucker, B. V. (2022). Temporal and spectral characteristics of conversational versus read fricatives in American English. *The Journal of the Acoustical Society of America*, *152*(4), 2073-2081.

Krashen, S. (1981). Second language acquisition. *Second Language Learning*, *3*(7), 19-39.

Kuznetsova A., Brockhoff P.B. and Christensen R.H.B. (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software*, 82(13), pp. 1–26. doi: 10.18637/jss.v082.i13.

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, *2*(12), 59-67.

Li, M. (2019). On the negative transfer of Chinese Northern Dialect to English Phonetics and the teaching strategies. *Foreign Language Education & Research*, 2019(02), 11-18. doi:10.16739/j.cnki.cn21-9203/g4.2019.02.002.

Liu, N. (刘乃华)(1988).汉英语音系统主要特点之比较. *南京师大学报(社会科学版)*(03),79-84. doi:CNKI:SUN:NJSS.0.1988-03-018.

Liu, S. (刘世生)(1990).英语语音教学浅探——汉语方音对英语语音的影响. *山东外语教学*(04),8-11. doi:10.16482/j.sdwy37-1026.1990.04.005.

R Core Team (2022). "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing (Vienna, Austria), version 4.1.2, https://www.R-project.org

Richards. J, Platt. J, & Weber. H. (1985). Longman Dictionary of Applied Linguistics. England: Longman.

San, D. (2007). *The phonology of standard Chinese*. OUP Oxford.

Shadle, C. H., & Mair, S. J. (1996, October). Quantifying spectral characteristics of fricatives. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96* (Vol. 3, pp. 1521-1524). IEEE.

VanPatten, B., Keating, G. D., & Wulff, S. (Eds.). (2020). *Theories in second language acquisition: An introduction*. Routledge.

Yue, M., Ling, D. (1994).汉语各方言区学生英语发音常误分析──汉英语音对比系列研究（之三）. *外语研究*(03). doi:CNKI:SUN:NWYJ.0.1994-03-012.

APPENDIX A

P-values for all examined parameters

Spectral mean

	f	v	θ	ð	s	z	ʃ	ʒ	h
onset	0.961	0.118	0.286	0.903	0.84	**0.017**	0.353	*0.061	0.934
midpoint	0.918	**0.023**	0.133	0.186	0.359	*0.059	0.564	0.119	0.011
offset	0.071	0.725	**0.001**	*0.054	0.642	0.475	0.809	0.232	0.462

Spectral variance

	f	v	θ	ð	s	z	ʃ	ʒ	h
onset	**0.045**	0.473	0.101	**0.004**	*0.052	**0.000**	**0.017**	**0.002**	0.802
midpoint	0.467	0.545	0.063	**0.011**	**0.008**	**0.000**	**0.002**	**0.000**	0.203
offset	0.341	0.254	**0.027**	0.148	0.080	**0.000**	**0.002**	**0.006**	0.581

Spectral skewness

	f	v	θ	ð	s	z	ʃ	ʒ	h
onset	0.817	0.165	0.356	0.726	0.865	**0.0048**	0.623	**0.037**	0.778
midpoint	0.614	**0.028**	0.267	0.256	0.492	0.094	0.308	0.416	0.746
offset	**0.033**	0.982	**0.002**	*0.059	0.619	0.343	0.692	0.297	0.463

Spectral kurtosis

	f	v	θ	ð	s	z	ʃ	ʒ	h
onset	**0.003**	0.458	0.161	**0.003**	0.140	**0.000**	0.701	0.752	0.169
midpoint	0.668	0.073	0.109	**0.004**	**0.013**	**0.000**	**0.000**	0.604	**0.012**
offset	**0.035**	0.337	0.544	0.447	0.096	**0.027**	0.145	0.711	0.567

Spectral peak

	f	v	θ	ð	s	z	ʃ	ʒ	h
peak	0.624	**0.016**	0.591	0.743	0.467	0.939	0.895	0.648	0.645

Segment duration

	f	v	θ	ð	s	z	ʃ	ʒ	h
duration	0.541	0.112	0.253	0.932	0.374	**0.032**	0.384	0.782	**0.039**

APPENDIX B

P-values for all examined second language acquisition factors

Attention

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Midpoint spectral mean	NA	0.598	0.069	0.760	0.461
Midpoint spectral variance	NA	0.187	**0.023**	*0.062	0.171
Midpoint skewness	NA	0.837	0.093	0.826	0.493
Midpoint kurtosis	NA	0.530	*0.055	0.119	0.232
Spectral peak	NA	0.899	**0.040**	0.803	0.399
Duration	NA	0.659	0.097	0.167	0.691

Motivation

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Midpoint spectral mean	NA	NA	0.731	*0.065	0.854
Midpoint spectral variance	NA	NA	**0.022**	**0.009**	0.274
Midpoint skewness	NA	NA	0.938	0.149	0.910
Midpoint kurtosis	NA	NA	0.113	0.152	0.189
Spectral peak	NA	NA	0.932	0.216	0.931
Duration	NA	NA	0.509	0.577	0.805

Satisfaction

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Midpoint spectral mean	NA	0.468	0.087	0.805	NA
Midpoint spectral variance	NA	0.399	*0.052	**0.010**	NA
Midpoint skewness	NA	0.293	**0.044**	0.877	NA
Midpoint kurtosis	NA	0.975	0.063	*0.061	NA
Spectral peak	NA	0.420	*0.056	0.566	NA
Duration	NA	0.681	0.188	**0.047**	NA

Self-rated proficiency

	1	2	3	4	5
Midpoint spectral mean	NA	0.760	0.135	0.671	0.548
Midpoint spectral variance	NA	0.140	0.092	**0.0143**	0.086
Midpoint skewness	NA	0.722	0.137	0.515	0.797
Midpoint kurtosis	NA	0.586	0.149	0.086	*0.060
Spectral peak	NA	0.617	**0.042**	0.162	0.415
Duration	NA	0.832	0.924	0.549	0.946

Frequency of speaking English

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Midpoint spectral mean	NA	0.919	0.688	0.139	0.787
Midpoint spectral variance	NA	0.833	**0.043**	0.091	**0.018**
Midpoint skewness	NA	0.661	0.928	0.132	0.954
Midpoint kurtosis	NA	0.499	0.409	0.132	**0.029**
Spectral peak	NA	0.996	0.941	0.144	0.930
Duration	NA	0.919	0.444	0.778	0.808

Length of living in Canada (in years)

	0	1	3	4	5	6	8
Midpoint spectral mean	0.928	0.085	0.278	0.212	0.102	0.888	0.724
Midpoint spectral variance	0.576	0.168	0.856	**0.014**	*0.059	*0.057	0.169
Midpoint skewness	0.762	0.132	0.129	0.301	0.165	0.908	0.448
Midpoint kurtosis	0.482	0.360	0.827	0.236	0.315	**0.025**	0.076
Spectral peak	0.764	0.095	0.380	0.915	0.164	0.817	0.900
Duration	0.685	0.150	0.096	*0.057	**0.022**	0.990	0.407

APPENDIX C

Syllables and words with target sounds

Syllables:
asha ithi afa usu izi ava izhi asa ihi ufu adha ishi uzhu uhu atha ivi ushu isi uzu ifi aza uvu aha azha udhu idhi uthu

F: fall fine fight often waffle office proof cough chief
V: vein vice vacuum advance seven travel prove glove sleeve
Th: think thought thigh lengthen toothpaste wealthy myth cloth math
Dh: they though thus weather other brother bathe smooth breathe
S: sun sight sphere assess awesome insight glass voice lettuce
Z: zebra zip zone realization wizard noisy freeze cheese these
Sh: show shut shine initial lotion direction English wash brush
Zh: genre usual measure visual illusion vision (massage concierge) pleasure
H: have high huge unhappy downhill behave

APPENDIX D

Questionnaire for native Mandarin speakers

Level of education (which year are you in?) ___

Your age is ___ (years old).

Your biological sex is:
__ female __ male

How long have you lived in Canada? (In years, if in months please indicate by m, e.g., 8m for 8 months) _____

Have you lived in any other English-speaking countries? __No __Yes (and where? _____, for how long? _____)

When did you start to learn English? (In age) _____ year(s) old

In general, how often do you speak in English daily?
__Not at all __Rarely __Sometimes __Often __Always

In general, how often do you read in English?
__Not at all __Rarely __Sometimes __Often __Always

In general, how often do you write in English?
__Not at all __Rarely __Sometimes __Often __Always

In general, you would pay attention to your pronunciation when you speak English.
__ Strongly disagree __Disagree __Neutral __Agree __Strongly agree

In general, you are satisfied with your English pronunciation.
__ Strongly disagree __Disagree __Neutral __Agree __Strongly agree

Do you agree that conveying your idea clearly is more important than using proper English pronunciation?
__ Strongly disagree __Disagree __Neutral __Agree __Strongly agree

You would want to have a native-like English pronunciation and/or accent.
__ Strongly disagree __Disagree __Neutral __Agree __Strongly agree

How would you rate your proficiency in English? ___
(1 - not at all, 5 – very proficient)