

# Acquisition of non-sibilant anterior English fricatives by adult second language learners

Seth Wiener, Zhe Gao, Xiaomeng Li, and Zhiyi Wu  
Carnegie Mellon University

This study examined the acquisition of the non-sibilant anterior fricatives /v, θ, ð/ by adult second language (L2) English talkers. Twenty-four Mandarin Chinese-L2 English talkers read aloud fricative-initial words. These talkers were chosen as their L1 contained /f/ but not /v, θ, ð/. Twenty L1 English listeners were asked to identify the L2-produced speech and rate the talker's accent. On average, 69% of the fricatives were correctly identified. /v/ was the most difficult to correctly identify and was identified less accurately than /θ/ and /ð/. A 'moderate' accent was reported by L1 listeners, but accent rating did not predict L1 identification behavior. An exploratory acoustic analysis involving the correctly identified words from 22 talkers revealed that L2 talkers produced mean temporal differences used for voicing in line with published L1 data. Non-sibilant fricatives – particularly /v/ – may require pedagogical interventions to push L2 talkers off their learning plateau.

**Keywords:** fricatives, speech production, accent, acoustic phonetics

## 1. Introduction

Fricatives represent a unique challenge in language acquisition. They are acoustically and aerodynamically complex sounds, which require speakers to push a rapid, turbulent flow of air through the mouth (Shadle, 1990; Stevens, 1971). Given the fine motor control needed to accurately produce fricatives, it is unsurprising that fricatives are often among the last speech sounds children acquire in their first language (L1: Holliday, Reidy, Beckman, & Edwards, 2015; Li, Edwards, & Beckman, 2009; Moskowitz, 1975; Nissen & Fox, 2005; Nittrouer, 1995; Nittrouer, Studdert-Kennedy, & McGowan, 1989). As an example, evidence from a large, cross-sectional study of American child speech indicates that by age 5 only 63.5 percent of children correctly produce word-initial /θ/ (Smit, Hand, Freilinger, Bernthal, & Bird, 1990).

Adults similarly struggle to acquire fricatives in a second language (L2), though this difficulty stems not from a lack of motor control, but rather from the need to articulate new L2 fricatives that differ in voicing and/or place of articulation from that of L1 fricatives. For most adult learners, acquisition of L2 fricatives within a classroom context typically yields limited results and can take years – if not decades – of practice to approach native-like pronunciation (e.g., Lombardi, 2003; Lord, 2005; Rau, Chang, & Tarone, 2009; Schmidt & Meyers, 1995). Evidence from L1 Mandarin Chinese-L2 English adult talkers with nearly 20 years of English experience indicates that only 64 percent of /θ/ utterances are correctly produced (Huang & Evanini, 2016; Rau et al., 2009).

In this study, we examine the acquisition of non-sibilant fricatives by adult L1 Mandarin Chinese-L2 English talkers. These talkers were chosen because they have a rich L1 sibilant fricative inventory but not /v, θ, ð/. Because one aim of L2 acquisition involves being understood by native speakers, we present the L2 recordings to L1 English listeners and ask them to identify the perceived fricative-initial words and rate the perceived accent of the talkers. We additionally carry out an exploratory acoustic analysis on the correctly identified fricatives and discuss these acoustic findings in terms of established L1 English findings.

### 1.1 Acquisition of English fricatives by L1 Mandarin talkers

Beijing Mandarin (hereafter, ‘Mandarin’) has a five-way fricative contrast determined by place of articulation: labiodental /f/, alveolar /s/, post-alveolar /ʃ/, (alveolo-)palatal /ç/, and velar /x/ (Lee & Zee, 2003). Importantly, the Mandarin fricative inventory is different from that of American English in that Mandarin fricatives are all voiceless and none are produced as an (inter)dental (Duanmu, 2007; Lin, 2007). American English has a nine-way fricative contrast distinguished by both voicing and place of articulation (voiced phoneme listed second in pair): labiodental /f, v/, (inter)dental /θ, ð/, alveolar /s, z/, palato-alveolar /ʃ, ʒ/, and glottal /h/ (Ladefoged & Maddieson, 1996; Ladefoged & Johnson, 2014).

For an L1 Mandarin-L2 English learner, accurate production of the non-sibilant fricatives /v, θ, ð/ involves mastering voicing for /v/, a new place of articulation for /θ/, and the combination of voicing and place of articulation for /ð/. Accurate production of these speech sounds can be challenging for L2 learners and often results in substitution of another L1 phoneme, such as /f/ or /s/ for /θ/ (Hansen, 2001; Huang & Evanini, 2016; Rogers & Dalby, 2005; Zhang & Xiao, 2014; Zheng & Samuel, 2017).

As an example, Rau et al. (2009) used a sociolinguistic interview to collect formal readings and informal free conversation from L2 English talkers from China and Taiwan. Using a variationist approach (e.g., Cedergren & Sankoff,

1974), Rau et al. found that accurate productions of /θ/ were estimated at 70% for talkers from China and 76% for talkers from Taiwan. The authors concluded that /s/ substitution was driven, in part, by the phonetic environment: /θ/ was produced more accurately when it was followed by a low front vowel ('thank') or high front vowel ('think') and less accurately when it was followed by a low back vowel ('thunder') or in a consonant cluster ('three'). Speech style also affected accurate fricative production such that clear, formal speech elicited through reading word lists and passages involved less /s/ substitution than informal, conversational speech elicited through interviews and stories (see Maniwa, Jongman, & Wade, 2009 for similar L1 findings). Rau et al. also tentatively concluded that lexical frequency affected /s/ substitution: high-frequency words were more likely to be produced with /θ/. The authors, however, acknowledged that their design was not balanced and their results may have been driven by the high-frequency lexeme 'think' and its various forms.

Here we build on previous studies to examine to what degree production of these novel fricatives contributes to perceived accent. We follow Munro and Derwing (Derwing & Munro, 1997; Munro & Derwing, 1995, 1999, 2001) and consider accent as the deviation from an expected pronunciation pattern or norm. Crucially, this body of research has found that accentedness has little to no effect on native listeners' understanding of L2 speech: utterances that are rated as being highly accented can still be perfectly transcribed by L1 listeners (Munro & Derwing, 1995). Although the field has expanded to focus more on intelligibility (i.e., understanding) and comprehensibility issues (i.e., the effort involved in understanding; see Levis, 2005 for discussion), there has been a resurgence in L2 accent studies rooted in phonetic theory. Chief among these studies is work on L2 voice-onset-time acquisition across languages like Korean (e.g., Chang, 2012; Holliday, 2015), Spanish (e.g., Nagle, 2019; Schoonmaker-Gates, 2015; Schuhmann & Huffman, 2019), and Japanese (e.g., Vaughn, Baese-Berk, & Idemaru, 2019). Our study contributes to this research by examining accented L2 English fricatives. The first aim of this study is to therefore record L1 Mandarin-L2 English learners' productions of fricative-initial minimal pairs, such as 'fat' and 'that', and examine correct identification and accent ratings by L1 English listeners.

The second aim of this study is to carry out an exploratory acoustic analysis on the utterances that L1 listeners correctly identified. Whereas L1 Mandarin-L2 English talkers' difficulty producing non-sibilant fricatives is well documented, previous studies have yet to fully examine the acoustic characteristics of L2 learners' productions, particularly /v/ and /ð/ utterances. Simultaneous voicing and frication require a balance between articulatory configuration and aerodynamics and may result in unique challenges for /v/ and /ð/ (Ohala, 1983; see also Bjorndahl, 2018). Because we did not collect L1 recordings for our study (and in-

lab recording has been paused due to COVID-19), we cannot compare our L2 results to an L1 baseline group. Instead, we use previously established L1 findings (e.g., Forrest, Weismer, Milenkovic, & Dougall 1988; Jesus & Shadle, 2002; Jongman, Wayland, & Wong, 2000; Maniwa, Jongman, & Wade, 2008; Nittrouer, 2002; Shadle & Mair, 1996; Shadle, Mair, & Carter, 1996) to guide our exploratory analysis.

## 1.2 Acoustic characteristics of English fricatives

We examine a total of six acoustic properties of the speech signal relevant to English fricatives: amplitude (2), duration (2), and spectral properties (2). First, we examine amplitude, which serves as a measure of a fricative's energy or 'loudness.' Previous research on L1 English talkers has examined fricative amplitude in terms of its overall noise amplitude (e.g., Behrens & Blumstein, 1988; Stevens, 1960) and its amplitude relative to the vowel (e.g., Hedrick & Ohde, 1993; Stevens, 1985). These studies have shown that sibilants, labiodentals, and voiceless fricatives have greater overall noise amplitude and relative amplitude compared to non-sibilants, (inter)dentals, and voiced fricatives. We therefore examine a "normalized amplitude" of each fricative's entire noise portion subtracted from its neighboring vowel amplitude (following Behrens & Blumstein, 1988). We also examine a "relative amplitude" of F5 at the center of the fricative minus F5 at the vowel onset (following Hedrick & Ohde, 1993; see also Jongman et al., 2000).

Second, we examine the duration of the frication noise, which serves as a salient property of English fricative voicing distinctions. Voiced fricatives have shorter noise durations than voiceless fricatives and shorter durations relative to whole word durations than voiceless fricatives (Behrens & Blumstein, 1988a; Jongman, 1989). Whereas duration can distinguish some places of articulation in L1 English talkers, it is not a reliable property for distinguishing labiodentals from (inter)dentals (Jongman et al., 2000). Thus, we measure fricative duration and fricative-to-word relative duration (hereafter 'normalized duration').

Third, we examine two spectral properties of the frication noise. Because the size and shape of the oral cavity can influence the overall spectral shape of the fricative (Ladefoged & Johnson, 2014; Stevens, 1971, 1998; Stevens, 1960), the location of the spectral peak in the frication noise can distinguish sibilants from non-sibilants, voiceless from voiced fricatives, as well as labiodentals from (inter)dentals as produced by L1 English talkers (Jongman et al., 2000; Maniwa et al., 2009). Fricatives are also characterized by their statistical distribution over the frequency domain. These summary statistics or 'spectral moments' also serve to distinguish English fricatives in L1 talkers (Jongman et al., 2000; Maniwa et al., 2008, 2009; Nittrouer et al., 1989). Here we examine spectral mean (here-

after ‘spectral centroid’), which captures the fricative’s average frequency of the spectrum, weighted by the energy, and has been shown to distinguish voiceless (higher centroid) from voiced fricatives (lower centroid), but not labiodentals from (inter)dentals in L1 English talkers.

Like research into the production of fricatives, research into the perception of fricatives has primarily focused on L1 listeners perceiving native speech. Previous fricative perception studies have tested natural (Zeng & Turner, 1990), synthetic (Heinz & Stevens, 1961) and hybrid speech (Nitttrouer, 1995, 2002). This body of research showed that (1) spectral properties are helpful for distinguishing sibilants; (2) formant transition properties are helpful for distinguishing non-sibilants; (3) when spectral properties are ambiguous, formant properties take on more perceptual weight; (4) noise duration, amplitude, and glottal vibration are all helpful for distinguishing voicing (Cole & Copper, 1975; Harris, 1958; Hedrick & Ohde, 1993; Jongman et al., 2000; Stevens, Blumstein, Glicksman, Burton, & Kurowski, 1992).

To summarize, in this study we record L2 English talkers producing fricative-initial words. We play this L2 speech to L1 English listeners and ask them to identify the intended word and rate the perceived accent of the talker. This allows us to examine L1 listener behavior and whether perceived accentedness predicts this behavior. We also carry out an exploratory acoustic analysis on the correctly identified fricatives and explore six acoustic properties associated with voicing and place of articulation differences in English L1 talkers.

## 2. Materials and methods

### 2.1 Participants

Twenty-four L1 Mandarin-L2 English adults took part in the production task. Table 1 summarizes the participants’ background from their Language Experience and Proficiency Questionnaire responses (Marian, Blumenfeld, & Kaushanskaya, 2007). All participants were born in China but were living in the U.S. at the time of testing. No participant spoke an additional language or non-Mandarin dialect containing the target fricatives /v, θ, ð/.

An additional 20 participants (mean age = 21.2) took part in the listening identification task. These 20 participants were native speakers of American English who self-reported regular daily interactions with L2 talkers (mean L2 interaction rate = 25 percent; range: 10–80 percent; standard deviation = 20 percent). All 44 participants across both tasks were undergraduate or graduate students affiliated with the authors’ university and had normal speech and hearing. All participants

Table 1. L1 Mandarin-L2 English participant summary

|  | Mean (SD)  |
|--|------------|
| Age in years   | 28.2 (6.9) |
| Biological sex   | 12 F; 12 M |
| Age started L2 English learning in years                     | 8.7 (3.8)  |
| Length of residence in English-speaking environment in years | 1.5 (1.5)  |
| Self-rated English speaking [0: none – 10: perfect]          | 4.8 (1.9)  |
| Self-rated English accent [0: none – 10: pervasive]          | 4.7 (1.6)  |

gave written consent to participate in the study and were either paid or volunteered for the study.

2.2 Materials

Twenty-four /v, θ, ð/ initial English words (eight per phoneme) and 24 /f/ initial English words were first identified (mean word frequency per million = 28.83; Brysbaert & New, 2009). This resulted in 24 fricative-initial minimal pairs (e.g., foul-vowel). An additional 48 filler minimal pairs were created containing plosive-initial words (e.g. /d/-/t/, dip-tip) for a total of 96 words. Of the possible 48 fricative-initial words, we report on only 16 critical words. These 16 words were chosen for a lack of initial consonant clusters and high lexical frequency (mean frequency per million = 607.4; Brysbaert & New, 2009). These 16 words appeared before seven different vowels /ε, eɪ, ju, i, ɪ, ə, ɔ/. See OpenScience for all materials: <https://osf.io/zhbgk/>

2.3 Procedure

2.3.1 L2 production task

All tasks involved in the study protocol were approved by the authors’ institute’s committee on human research. As part of a larger pedagogical study on L2 English speech sound acquisition, participants performed the production task twice, exactly one week apart. Participants were seated in a sound-attenuated booth and instructed to read aloud the words as clearly and accurately as possible into a microphone approximately 4–6 inches away from their mouths. Speech was recorded at 16-bit/44.1 kHz using Praat (Boersma & Weenink, 2019). Each word was displayed on a computer monitor in a pseudo-randomized order such that the word-initial phoneme in consecutive trials always differed in manner and place of articulation and minimal pairs were separated by multiple trials. To

encourage more natural speaking rates, each word was shown for 1-second followed by a 1.5-second inter-trial-interval. Stimuli were presented using E-prime (version 2.0; Psychology Software Tools, 2007). Instructions were given in English and Mandarin by a bilingual experimenter. The production task lasted approximately 5–10 minutes.

Because each L2 participant produced the 16 critical targets twice, this resulted in 32 tokens per talker or 768 total tokens. From these tokens, 51 were removed due to experimenter recording errors or excessive participant noise (e.g., coughing, cell phones, etc.), resulting in 717 unique tokens. Figure 1 (top left) summarizes the L2 production task.

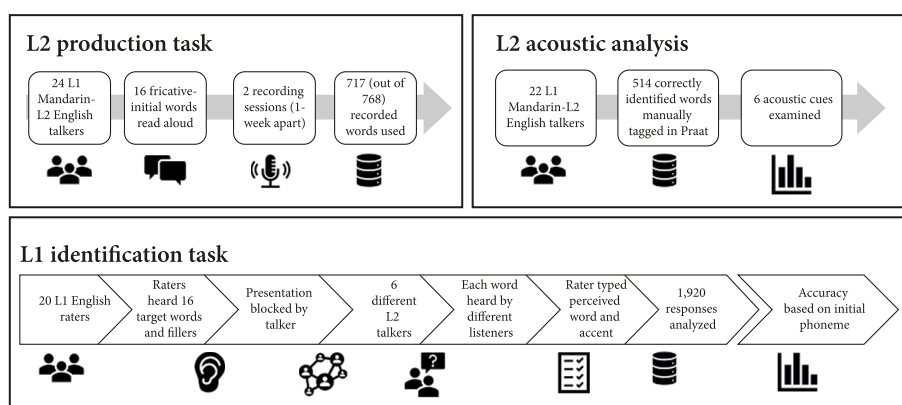


Figure 1. Study design

### 2.3.2 L1 identification task

L1 English listeners were seated in front of a computer in a quiet room and given headphones. Participants were told they would listen to L1 and L2 speech by several different talkers, but were not made explicitly aware of the target fricatives. Participants heard each word only once via headphones and were instructed to type the perceived English word. Participants were told to guess if they were unsure. The 717 L2 fricative-initial productions (along with 336 L2 plosive-initial fillers and 75 non-fricative-initial fillers produced by five L1 talkers) were spread among eight presentation lists, such that each list contained productions from only six of the 24 L2 talkers, and each L2 talker was included in at least two lists. Each list was divided into 12 blocks, with each block containing productions from only a single talker. The 12 blocks were comprised of seven L2 blocks (which included one practice block) and five L1 talker blocks. Each L2 block contained 30 pseudo-randomized trials (16 /f, v, θ, ð/-initial targets; 14 plosive-initial fillers), and each L1 talker block contained 15 randomized trials with no

fricative-initial words. These L1 talker blocks' primary purpose was for the larger two-week pedagogy study. Stimuli were presented using E-prime with a 1-second inter-trial interval.

At the end of each L1 and L2 talker block, participants were asked to rate the talker's accent from 0 to 10 (0: None; 1: Almost none; 2: Very light; 3: Light; 4: Some; 5: Moderate; 6: Considerable; 7: Heavy; 8: Very heavy; 9: Extremely heavy; 10: Pervasive). The options were presented with both the index number and the corresponding accent level in English. The first block was treated as a practice block, after which listeners could ask the experimenter any questions. In total, the 20 L1 English listeners responded to 16 fricative-initial targets per block across six L2 blocks for a total of 1,920 responses. The 20 L1 listeners each provided six L2 accent ratings (1 per talker  $\times$  6 L2 talker blocks) for 119 total ratings (one response was lost). The identification task lasted approximately 20 minutes. Figure 1 (bottom) summarizes the L1 identification task.

### 2.3.3 *Acoustic analysis*

The 514 tokens in which the onset was correctly identified by at least one L1 listener were analyzed using Praat version 6.1.08 (Boersma & Weenink, 2019). Note that two participants' recordings were removed entirely due to their poor recording quality. Analyses involved simultaneous consideration of the spectrogram and waveform. "Normalized amplitude" captured the root-mean-square (rms) amplitude (in dB) of each fricative's entire noise portion. This amplitude was then subtracted from vowel amplitude, which consisted of rms amplitude (in dB) averaged over three consecutive pitch periods during maximum vowel amplitude (following Behrens & Blumstein, 1988). "Relative amplitude" captured the difference between the amplitude (in dB) of F5 at the center of the fricative and amplitude of F5 at the vowel onset (following Hedrick & Ohde, 1993; see also Jongman et al., 2000).

Duration was calculated following Jongman et al. (2000). Fricative onset was defined as the nearest zero crossing at which high-frequency energy first appeared. Fricative offset was defined as the nearest zero crossing containing the intensity minimum before the onset of vowel periodicity (voiceless) or the nearest zero crossing at which pitch exhibited the earliest change in the waveform (voiced). Because words differed in their final phoneme, duration was defined as the interval between the fricative onset and the nearest zero crossing at which waveform and high-frequency energy ceased. Normalized duration was calculated for each token by taking the fricative duration and dividing it by the whole word duration.

Spectral peak was calculated using a modified Praat script by Styler (2014). Peak estimation was based on spectra generated using fast Fourier transform. A



20-ms Hamming window was used to calculate peaks at three points in the frication: one-third, midpoint, and two-thirds. Spectral peak was defined as the highest frequency measured across all three slices' bins.

Spectral centroid was calculated for each fricative using a modified Praat script by DiCanio (2013; see also DiCanio et al., 2020). The script used multiple discrete Fourier transforms across the duration of the fricative to arrive at time-averaged measures of the spectral moments (see Shadle, 2012). Six 15-ms windows were used with a low pass filter cut-off set to 300 Hz (see Maniwa et al., 2009). In total, six acoustic properties were examined. Figure 1 (top right) summarizes the acoustic analysis.

### 3. Results

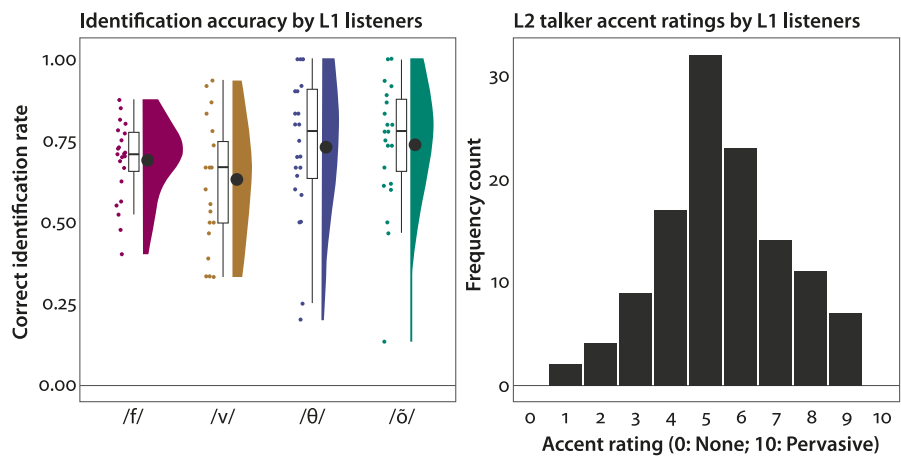
See OpenScience for data and R code detailing our statistical analyses: <https://osf.io/zhbgbk/>

#### 3.1 Identification of L2 fricative-initial words by L1 listeners

The 1,920 L1 listener responses were first analyzed for inter-rater reliability: the expected chance probability of agreement was subtracted from the observed probability of agreement, which was then divided by one minus the chance probability of agreement. The L1 listeners demonstrated moderate inter-rater agreement:  $\kappa = .47$  (see Landis & Koch, 1977).

Next, the overall average correct word identification was calculated: 50 percent of all responses were correctly identified as the intended word. However, because misperception of vowels and other non-fricative consonants (e.g., reporting 'fair' for 'fail') was irrelevant to the present research question, each response was manually scored as either having the correctly intended initial fricative (1) or an incorrect initial consonant (0). Because the L2 participants showed no production accuracy difference between the two lab visits ( $t(21) = 1.19, p = .25$ ), the production data were combined to increase statistical power. The initial fricative was correctly identified across 69 percent of all responses. Voiceless (70 percent) and voiced (68 percent) fricatives were produced with similar mean accuracies whereas labiodentals (67 percent) were produced, on average, slightly less accurately than (inter)dentals (73 percent). Figure 2 (left) shows the overall results by target fricative using the raincloud plotting scheme (Allen et al., 2019). This plot shows individual means for each L2 talker (with jitter added), boxplots with medians and quantiles, fricative means (large black points), and overall density curves for each fricative. On average, talkers' accent was scored as 5.5 (95% confidence interval

[5.2, 5.8]), which was ‘moderate’ according to our scale. Figure 2 (right) shows a histogram of the accent responses indicating a fairly normal distribution across talkers.



**Figure 2.** L1 English listeners’ identification accuracy (left); Histogram of L1 listeners’ accent ratings (right)

Table 2 presents a confusion matrix of the target fricatives. This table indicates that /ð/ was identified most accurately at nearly 74 percent and /v/ identified least accurately at nearly 63 percent. /f/ was most often misheard as /v/ whereas /v/ was most often misheard as /w/. /θ/ was most often misheard as /f/ (but also /s/ to a relatively large degree) whereas /ð/ was most often misheard as /b/ and /f/.

**Table 2.** Confusion matrix of /f, v, θ, ð/ by reported initial consonant (%). Note: ‘None’ indicates a null consonant onset

|     | /p/ | /b/ | /m/ | /f/  | /v/  | /θ/  | /ð/  | /t/ | /d/ | /n/ | /s/ | /z/ | /ɹ/ | /l/ | /w/  | None |
|-----|-----|-----|-----|------|------|------|------|-----|-----|-----|-----|-----|-----|-----|------|------|
| /f/ | 0.4 | 3.1 | 0.1 | 69.0 | 10.4 | 6.4  | 7.5  | 1.1 | 0.3 | 0.0 | 0.7 | 0.0 | 0.1 | 0.0 | 0.2  | 0.6  |
| /v/ | 0.0 | 5.6 | 5.0 | 3.9  | 63.1 | 0.0  | 1.9  | 0.0 | 0.0 | 1.1 | 0.0 | 0.0 | 7.5 | 0.6 | 10.6 | 0.8  |
| /θ/ | 0.0 | 0.0 | 0.0 | 11.2 | 0.0  | 72.9 | 0.0  | 0.8 | 4.2 | 0.4 | 8.8 | 0.0 | 0.4 | 0.0 | 0.4  | 0.8  |
| /ð/ | 0.0 | 4.4 | 0.3 | 4.2  | 1.0  | 3.3  | 73.6 | 0.6 | 2.8 | 0.0 | 1.9 | 1.9 | 0.8 | 1.4 | 0.3  | 3.3  |

To analyze responses by target phoneme and perceived accent, a mixed-effects logistic regression model was built in R (version 3.6.2; R core team, 2019) using the *lme4* package (Bates et al., 2014). The dependent variable was accurate initial phoneme identification (coded as 1 or 0), with /f/ coded as the reference level allowing for three categorical contrasts: /f/-/v/, /f/-/θ/, and /f/-/ð/ (addi-

tional contrasts were obtained by releveling the model; see R code on OpenScience). Standardized ratings were calculated by subtracting the overall mean rating from the talker's rating and then dividing by the standard deviation. Standardization allowed for more interpretable results in our regression models. Neither talker block ( $\chi^2(1) = 2.50$ ,  $p = .11$ ) nor accent rating ( $\chi^2(1) = 0.03$ ,  $p = .86$ ) was included in the final model as these variables did not improve model fit. Random participant and item intercepts were included. Table 3 summarizes the model's output along with the R code.

**Table 3.** Mixed effects logistic regression model (correct fricative identification)

|                                  | $\beta$ estimate | SE   | Z     | p     |
|----------------------------------|------------------|------|-------|-------|
| (Intercept: /f/ reference level) | 1.22             | 0.16 | 7.76  | <.001 |
| /f/ – /v/                        | –0.40            | 0.23 | –1.73 | .08   |
| /f/ – /θ/                        | 0.34             | 0.24 | 1.43  | .15   |
| /f/ – /ð/                        | 0.29             | 0.28 | 1.02  | .31   |

R code: `glmer(onset.accuracy~phoneme + (1|rater)+(1|item), family="binomial", optimizer="bobyqa")`

The model identified differences among the four phonemes' log-odds of accurate identification: /ð/ had higher log-odds than /v/ ( $p < .01$ ), and /θ/ had higher log-odds than /v/ ( $p < .05$ ). /f/ did not have higher log-odds than /v/ ( $p = .08$ ). In other words, /f, θ, ð/ did not differ from one another in terms of their log-odds, whereas /v/ consistently had the lowest log-odds across all three comparisons.

In sum, L1 listeners correctly identified nearly 70 percent of the fricatives with /θ, ð/ identified statistically more accurately than /v/; /f/ and /v/ performance did not differ. Overall, L1 listeners found L2 talkers to have a moderate accent. This accent rating, however, did not predict word identification behavior.

### 3.2 Acoustic analysis of L2 fricatives

The exploratory acoustic analysis was carried out on the 514 tokens in which at least one L1 listener correctly identified the onset phoneme. Figure 3 uses the same raincloud plotting scheme (Allen et al., 2019) and shows the amplitude (top two plots), duration (middle two plots), and spectral properties (bottom two plots) of the fricatives.

To compare whether the six acoustic measurements statistically differed between place and voicing (and to test for a possible two-way interaction), analyses were carried out in R using the *lme4* package. For each model, the dependent variable was the relevant acoustic property; fixed effects included voicing

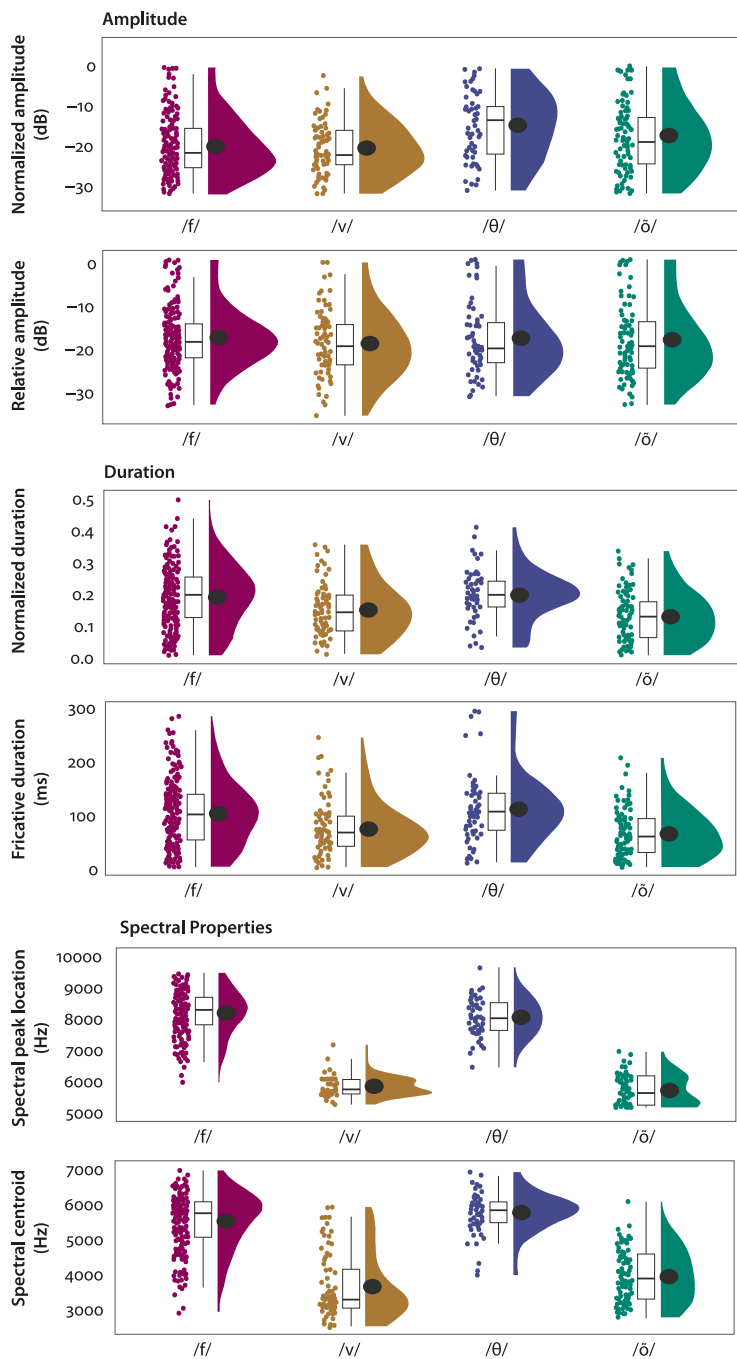


Figure 3. Raincloud plots for the six acoustic properties

(dummy coded with “voiced” as the reference level), place of articulation (dummy coded with “(inter)dental” as the reference level), and the two-way interaction. Talker and item random intercepts were included in each model. Because six different linear models were tested in total, reported *p*-values were adjusted using Bonferroni correction. All models’ residuals were found to be normally distributed. See R code on OpenScience for full output of all models.

Voiceless and voiced fricatives had similar mean normalized amplitudes (voiceless = -19.01 dB; voiced = -19.07 dB;  $p = .99$ ) and relative amplitudes (voiceless = -17.28 dB; voiced = -18.18 dB;  $p = .99$ ). Labiodental and (inter)dental fricatives had similar normalized amplitudes (labiodental = -20.22 dB; (inter)dental = -16.55 dB;  $p = .14$ ) and relative amplitudes (labiodental = -17.62 dB; (inter)dental = -17.63 dB;  $p = .99$ ). No two-way interaction was found ( $ps > .9$ ).

Words containing voiceless fricatives were, on average, longer than words containing voiced fricatives (voiceless words = 526 ms; voiced words = 497 ms). This duration difference was statistically significant in terms of fricative duration (voiceless = 106 ms; voiced = 71 ms;  $p = .02$ ) but not in terms of normalized duration (voiceless = .20; voiced = .14;  $p = .25$ ). Neither normalized nor fricative duration differences were observed between the labiodental and (inter)dental fricatives (labiodental fricative duration = 96 ms; (inter)dental fricative duration = 84 ms; labiodental normalized duration = .18; (inter)dental normalized duration = .16;  $ps > .4$ ). No two-way interaction was found ( $ps > .10$ ).

Voiceless fricatives showed higher spectral peaks than voiced fricatives (voiceless = 8202 Hz; voiced = 5789 Hz;  $p < .001$ ). Labiodental fricatives showed similar spectral peaks to (inter)dental fricatives (labiodental = 7594 Hz; (inter)dental = 6612 Hz;  $p = .32$ ). No two-way interaction was found ( $p = .99$ ). Voiceless fricatives showed higher spectral centroids than voiced fricatives (voiceless = 5584 Hz; voiced = 3839 Hz;  $p < .001$ ). Labiodental fricatives showed similar spectral centroids to (inter)dental fricatives (labiodental = 5036 Hz; (inter)dental = 4663 Hz;  $p = .23$ ). No two-way interaction was found ( $p = .99$ ).

To summarize the exploratory acoustic analysis, L2 talkers produced voiceless/voiced differences for fricative duration, spectral peak, and spectral centroid. No mean acoustic differences were observed for any amplitudinal measurement or any comparison involving place of articulation.

#### 4. Discussion

This study examined the acquisition of the anterior non-sibilant fricatives /v, θ, ð/ by L1 Mandarin-L2 English talkers. These L2 talkers were chosen because /v, θ, ð/ are all novel L2 speech sounds and notoriously difficult for adult learners

to master (Huang & Evanini, 2016; Lombardi, 2003; Lord, 2005; Rau et al. 2009; Schmidt & Meyers, 1995). We recorded L2 talkers reading aloud fricative-initial minimal pairs. L1 English listeners were asked to identify the words and rate the talkers' accents. L1 listeners identified /f, θ, ð/ all with comparable accuracies and all proportionally greater than /v/, which had the lowest mean identification accuracy. Identification accuracy of /v/ was significantly lower than identification accuracy of /θ/ and /ð/ (but not /f/).

Our identification results suggest that L2 talkers often substituted a voiceless fricative from their L1, like /f/ or /s/ for the intended non-native sound (e.g., Hansen, 2001; Zhang & Xiao, 2014). Table 2 showed that a wide range of other speech sounds were reported by L1 listeners, whose inter-rater agreement was only moderate. /f/ was perceived most often as /v/ (10.4%; primarily driven by the high frequency 'very' responses),<sup>1</sup> suggesting a lack of clear voicing cues for these two labiodentals. /θ/ was perceived most often as /f/ (11.2%) but also as /s/ (8.8%) to a fairly high extent. This corroborates previous studies that reported substitution of another L1 phoneme for /θ/ (e.g., Rau et al., 2009). /ð/ was perceived most often as /b/ (4.4%), which was presumably another frequency-driven error given the 'been' responses. /v/ was perceived most often as /w/ (10.6%) due to 'when' and 'where' responses, which could be a combination of frequency and hypercorrection when the talker has the labiodental approximant /v/ in their L1 Mandarin dialect. Beijing and northern Mandarin talkers are noted for producing /v/ rather than /w/ in their speech (Shen, 1987; Wiener & Shih, 2013). Under this account, /v/ was used in the present study in place of novel phonemes like /v, ð/. As a result, /ðer/ became /vɛr/ and /ðɪn/ became /vɪn/, which L1 listeners most likely perceived as 'where' and 'when.' Whereas three of our participants reported speaking an additional non-Mandarin dialect (Shanghai, Xiangyang, and Li dialects), we did not ask participants to further specify whether they spoke Mandarin with a Beijing or northern accent. Future studies may explore our hypothesis with a more controlled participant population.

Regarding accent, our L1 listeners reported comparable mean accent ratings (5.5) to those self-reported by the L2 talkers (4.8). Our talkers and listeners were, therefore, in agreement over what "accent" represented and both groups agreed that the L2 talkers had a 'moderate' accent. Yet, we found that accent ratings did not predict overall identification for any phoneme. These findings corroborate previous results that showed perceived accentedness has a limited (if any) effect

---

1. It is unclear how the lexical frequency of the intended word and the lexical frequency of the perceived word affected the results. For example, if a talker had intended to say 'ferry' but it was perceived as 'very' (which is higher in frequency), lexical frequency may have actually caused misperception. We include frequency information on OpenScience for readers to explore.

on L1 listeners' ability to correctly identify speech (e.g., Munro & Derwing, 1995, 2001).

Whether accent was linked to the phonetics of the fricatives, to the competition among phonological categories (particularly those that exist in both the L1 and L2), or some combination of phonetics-phonology is an open question. Unfortunately, our filler plosive trials may have unintentionally contributed to the overall accent ratings at some level. Whereas our L1 listeners correctly identified an average of 84 percent of the filler L2 plosive initials – a 15 percent mean increase from the L2 fricative initials – L1 listeners still struggled to correctly identify the intended phonetic category. The presence of plosive-initial words appears to have influenced fricative perception, as suggested by the plosive-initial misperceptions shown in the confusion matrix (Table 2; see also Xie et al., 2017; Xie & Myers, 2017 for L1 Mandarin-L2 English plosive results). As the field of L2 pronunciation research advances, particularly with the increase in rigorous internet-based studies (e.g., Nagle & Huensch, 2020), the consideration of fillers is increasingly important: what role should they play (if any) in experimental L2 pronunciation studies?

We also conducted an exploratory acoustic analysis on the correct L2 productions. We found that L2 talkers produced mean frication noise, spectral peak, and spectral centroid differences for voiceless /f, θ/ compared to voiced /v, ð/. Using Jongman et al.'s (2000) L1 English results as an *estimated* L1 baseline, we observed relatively similar fricative durations (cf. Jongman et al., Table 6: /f, θ/ = 164.5 ms; /v, ð/ = 84 ms) and relatively similar spectral peaks (cf. Jongman et al., Figure 1: /f, θ/ ≈ 8000 Hz; /v, ð/ ≈ 7000 Hz). Jongman et al. did not report spectral centroid for voiceless and voiced non-sibilants separately.

To what degree are our results comparable to L1 English results like those of Jongman et al. (2000)? We believe the duration finding is reliable as both studies recorded and analyzed consonant-vowel fricative initial-syllables/words. The method of measurement was identical and the overall means between the two studies are relatively similar in that voiceless /f, θ/ were longer in duration than voiced /v, ð/. This finding is not entirely unexpected given that adults already familiar with fricative production should transfer their ability to produce turbulent airflows to another language. Moreover, sustaining frication appears to be a relatively similar phonation process across languages (Gordon, Barthmaier, & Sands, 2002). Our finding that L1 Mandarin-L2 English talkers can produce non-native durational properties of a novel voiced fricative corroborates previous experimental studies on other L2 phonetic temporal contrasts (e.g., Arslan & Hansen, 1997; Flege, 1993), though we know of no such study specifically targeting L2 fricatives. Additionally, like L1 talkers reported in Jongman et al. (2000), L2 talkers only produced mean temporal differences for voicing. No statistical difference was seen for

place of articulation. We note that our population of L2 talkers did not show a statistical difference for normalized duration, despite the .06 mean difference (voiceless = .20; voiced = .14). We believe this is a power issue given our relatively small sample size. A statistical difference was found ( $p = .04$ ) before applying Bonferroni correction (see our R code on OpenScience for details).

We did not observe a difference in amplitudinal properties for either voicing or place of articulation. We followed Jongman et al.'s (2000) methods, though the two studies' vowels and recording procedures differed slightly. It is unclear if this null finding is because amplitudinal properties are relatively difficult to acquire in an L2 or because our recordings were too dissimilar from those of Jongman et al.'s. Given that all Mandarin fricatives are voiceless, it seems likely that our population generally produced the English anterior fricatives with greater intensity than L1 English talkers typically do (e.g., Huang & Evanini, 2016; Rau et al., 2009). Future studies may explore this question with a more carefully controlled design similar to Maniwa et al.'s (2009).

We present our L2 spectral peak and centroid results as *exploratory* findings for L1-like differences in spectral cues relevant for voiceless /f, θ/ compared to voiced /v, ð/. We refrain from further interpretation of the spectral results. Differences in sampling rate, window size, and window placement are all known to affect spectral moments (see Shadle, 1990, 2012) and ultimately make our results too dissimilar to those of Jongman et al. (2000) for any substantial comparison.

Taken together, our results paint a sobering picture for adult L2 anterior non-sibilant fricative acquisition. At the highest level, our L2 population produced words that were correctly identified by L1 listeners only 50 percent of the time and word-initial fricatives that were correctly identified only 69 percent of the time. These means are nearly identical to those of previous studies targeting L1 Mandarin-L2 English talkers (Huang & Evanini, 2016; Rau et al., 2009). Like Huang and Evanini's population, our sample population included L2 talkers who had an average of twenty years of L2 English study and were immersed in their L2 environment, yet still routinely produced speech in a manner that led to incorrect identification by L1 listeners. Simultaneous voicing and frication for /v/ appeared to be especially difficult for many of our L2 talkers. Of the three novel phonemes, /v/ lagged behind /θ/ and /ð/ in L1 identification accuracy. What made /v/ harder than /θ/ and /ð/ for many of our talkers (and listeners)? Our exploratory acoustic analysis indicated that there were no reliable place of articulation cues produced that might help listeners discriminate between these labiodental and (inter)dental fricatives. It is worth pointing out that /v/ is a relatively hard phoneme to produce and perceive (Ohala, 1983). Recent findings by Bjorndahl (2018) suggest that the cross-linguistic variability of /v/ in terms of realization and phonological patterning makes /v/ production a unique challenge in speech learning. Here we extend



these claims to adult L2 learners who lack /v/ in their L1. Finding the right balance between articulatory configuration and aerodynamics can be extremely difficult for adult L2 learners and often leads to an L2 learning plateau.

We conclude by noting two limitations to our study. First, our exploratory acoustic analysis included a relatively small number of recordings ( $N=514$ ). A much larger and more carefully controlled sample is necessary to fully understand the acoustic patterns of L2 fricatives. As discussed earlier, a proper L1 baseline group – rather than published means – is needed to better situate the L2 results. Second, an unintended consequence of our stimuli – only realized post-hoc thanks to Allard Jongman – was that our target /ð/ words are all function words (an idiosyncratic characteristic of English). Additionally, several of our items had additional minimal pairs we did not originally foresee (e.g., ‘sail’ in our ‘fail’ – ‘veil’ pair). Future studies should consider a cleaner design in which the listener only has to choose between a set number of words, i.e., a 2-alternative-forced-choice task. Additionally, some of our L2 participants may have been forced to speak faster than they would have normally spoken and none of our talkers knew what word was coming next. Both speaking style (Maniwa et al., 2009; Rau et al., 2009) and task (Declerck & Kormos, 2012) have been shown to affect speech production.

## 5. Conclusion

This study demonstrated that L1 Mandarin-L2 English talkers with extensive English experience often plateau in their English non-sibilant fricative production ability. Overall, our population of L2 talkers produced /f, v, ð, θ/ utterances that were correctly identified by L1 listeners roughly 70 percent of the time. /v/ was a particularly difficult novel sound for L2 talkers to acquire. L1 listeners reported that the L2 talkers had a ‘moderate’ accent, but this rating did not predict L1 listener identification accuracy. An exploratory acoustic analysis revealed that the L2 talkers produced temporal properties used to distinguish between voiceless and voiced fricatives in a relatively similar manner to that of reported L1 English talkers (cf. Jongman et al., 2000).

## Acknowledgements

The authors thank Joy Maa for help with the experiment, Will Styler and Christian DiCanio for help with Praat scripts, and Christina Bjorndahl for general help with fricatives. Jeff Holli-day, Yung-hsiang Shawn Chang, Allard Jongman, and three anonymous reviewers all provided incredibly valuable comments on earlier versions of the manuscript.

## References

- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. T., & Kievit, R. A. (2019). Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Research*, 4(63). <https://doi.org/10.12688/wellcomeopenres.15191.1>
- Arslan, L. M., & Hansen, J. H. (1997). A study of temporal features and frequency characteristics in American English foreign accent. *The Journal of the Acoustical Society of America*, 102(1), 28–40. <https://doi.org/10.1121/1.419608>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *ArXiv e-prints*, 1406.5823. <https://doi.org/10.18637/jss.v067.i01>
- Behrens, S., & Blumstein, S. E. (1988). On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants. *The Journal of the Acoustical Society of America*, 84(3), 861–867. <https://doi.org/10.1121/1.396655>
- Behrens, S. J., & Blumstein, S. E. (1988a). 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](https://doi.org/10.1016/S0095-4470(19)30504-2)
- Bjorndahl, C. (2018). A Story of /v/: Voiced Spirants in the Obstruent-Sonorant Divide (Doctoral dissertation). Retrieved from Cornell Thesis and Dissertation. <https://doi.org/10.7298/X4BZ648J>
- Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1.08, retrieved April 2019 from <http://www.praat.org/>
- Brysaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990. <https://doi.org/10.3758/BRM.41.4.977>
- Cedergren, H. J., & Sankoff, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language*, 50(2), 333–355. <https://doi.org/10.2307/412441>
- Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first language speech production. *Journal of Phonetics*, 40(2), 249–268. <https://doi.org/10.1016/j.wocn.2011.10.007>
- Cole, R. A., & Cooper, W. E. (1975). Perception of voicing in English affricates and fricatives. *The Journal of the Acoustic Society of America*, 58, 1280–1287. <https://doi.org/10.1121/1.380810>
- Declerck, M., & Kormos, J. (2012). The effect of dual task demands and proficiency on second language speech production. *Bilingualism: Language and Cognition*, 15(4), 782–796. <https://doi.org/10.1017/S1366728911000629>
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19, 1–16. <https://doi.org/10.1017/S0272263197001010>
- DiCanio, C. (2013). Spectral moments Praat script [code] [http://www.acsu.buffalo.edu/~cdicanio/scripts/Time\\_averaging\\_for\\_fricatives\\_2.o.praat](http://www.acsu.buffalo.edu/~cdicanio/scripts/Time_averaging_for_fricatives_2.o.praat)
- DiCanio, C., Zhang, C., Whalen, D. H., & García, R. C. (2020). Phonetic structure in Yoloxóchitl Mixtec consonants. *Journal of the International Phonetic Association*, 50(3), 333–365. <https://doi.org/10.1017/S0025100318000294>
- Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford: Oxford University Press.
- E-Prime [Computer Software]. (2007). Version 2.0. Pittsburgh: Psychology Software Tools.

- Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America*, 93(3), 1589–1608. <https://doi.org/10.1121/1.406818>
- 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. <https://doi.org/10.1121/1.396977>
- Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, 32(2), 141–174. <https://doi.org/10.1017/S0025100302001020>
- Hansen, J. G. (2001). Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics*, 22(3), 338–365. <https://doi.org/10.1093/applin/22.3.338>
- 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>
- Hedrick, M. S., & Ohde, R. N. (1993). Effect of relative amplitude of frication on perception of place of articulation. *The Journal of the Acoustical Society of America*, 94(4), 2005–2026. <https://doi.org/10.1121/1.407503>
- Heinz, J. M., & Stevens, K. N. (1961). On the properties of voiceless fricative consonants. *The Journal of the Acoustical Society of America*, 33(5), 589–596. <https://doi.org/10.1121/1.1908734>
- Holliday, J. J. (2015). A longitudinal study of the second language acquisition of a three-way stop contrast. *Journal of Phonetics*, 50, 1–14. <https://doi.org/10.1016/j.wocn.2015.01.004>
- Holliday, J. J., Reidy, P. F., Beckman, M. E., & Edwards, J. (2015). Quantifying the robustness of the English sibilant fricative contrast in children. *Journal of Speech, Language, and Hearing Research*, 58(3), 622–637. [https://doi.org/10.1044/2015\\_JSLHR-S-14-0090](https://doi.org/10.1044/2015_JSLHR-S-14-0090)
- Huang, B., & Evanini, K. (2016). Think, sink, and beyond: Phonetic variants and factors contributing to English th pronunciation among Chinese speakers. *Journal of Second Language Pronunciation*, 2(2), 253–275. <https://doi.org/10.1075/jslp.2.2.06hua>
- Jesus, L. M., & Shadle, C. H. (2002). A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics*, 30(3), 437–464. <https://doi.org/10.1006/jpho.2002.0169>
- Jongman, A. (1989). Duration of frication noise required for identification of English fricatives. *The 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), 1252–1263. <https://doi.org/10.1121/1.1288413>
- Ladefoged, P., & Johnson, K. (2014). *A course in phonetics* (7th ed.). Stamford, CT: Cengage Learning.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford Cambridge, MA: Blackwell.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Lee, W. S., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association*, 33(1), 109–112. <https://doi.org/10.1017/S0025100303001208>
- Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL quarterly*, 39(3), 369–377. <https://doi.org/10.2307/3588485>

- Li, F., Edwards, J., & Beckman, M.E. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics*, 37(1), 111–124. <https://doi.org/10.1016/j.wocn.2008.10.001>
- Lin, Y.H. (2007). *The Sounds of Chinese*. Cambridge: Cambridge University Press.
- Lombardi, L. (2003). Second language data and constraints on manner: Explaining substitutions for the English interdentals. *Second Language Research*, 19(3), 225–250. <https://doi.org/10.1177/026765830301900304>
- Lord, G. (2005). (How) can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania*, 88(3), 557–567. <https://doi.org/10.2307/20063159>
- Maniwa, K., Jongman, A., & Wade, T. (2008). Perception of clear fricatives by normal-hearing and simulated hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 123(2), 1114–1125. <https://doi.org/10.1121/1.2821966>
- 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>
- Marian, V., Blumenfeld, H.K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50(4), 940–967. [https://doi.org/10.1044/1092-4388\(2007/067\)](https://doi.org/10.1044/1092-4388(2007/067))
- Moskowitz, B.A. (1975). The acquisition of fricatives: A study in phonetics and phonology. *Journal of Phonetics*, 3(3), 141–150. [https://doi.org/10.1016/S0095-4470\(19\)31361-0](https://doi.org/10.1016/S0095-4470(19)31361-0)
- Munro, M.J., & Derwing, T.M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38(3), 289–306. <https://doi.org/10.1177/002383099503800305>
- Munro, M.J., & Derwing, T.M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49(Suppl 1), 285–310. <https://doi.org/10.1111/0023-8333.49.s1.8>
- Munro, M.J., & Derwing, T.M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, 451–468. <https://doi.org/10.1017/S0272263101004016>
- Nagle, C.L. (2019). A longitudinal study of voice onset time development in L2 Spanish stops. *Applied Linguistics*, 40(1), 86–107. <https://doi.org/10.1093/applin/amlx011>
- Nagle, C.L., & Huensch, A. (2020). Expanding the scope of L2 intelligibility research: Intelligibility, comprehensibility, and accentedness in L2 Spanish. *Journal of Second Language Pronunciation*. <https://doi.org/10.1075/jslp.20009.nag>
- Nissen, S.L., & Fox, R.A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *The Journal of the Acoustical Society of America*, 118(4), 2570–2578. <https://doi.org/10.1121/1.2010407>
- Nittrouer, S. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *The Journal of the Acoustical Society of America*, 97(1), 520–530. <https://doi.org/10.1121/1.412278>
- Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *The Journal of the Acoustical Society of America*, 112(2), 711–719. <https://doi.org/10.1121/1.1496082>

- Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech, Language, and Hearing Research*, 32(1), 120–132. <https://doi.org/10.1044/jshr.3201.120>
- Ohala, John J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 189–216). New York, NY: Springer. [https://doi.org/10.1007/978-1-4613-8202-7\\_9](https://doi.org/10.1007/978-1-4613-8202-7_9)
- Rogers, C. L., & Dalby, J. (2005). Forced-choice analysis of segmental production by Chinese-accented English speakers. *Journal of Speech, Language, and Hearing Research*, 48(2), 306–322. [https://doi.org/10.1044/1092-4388\(2005\)021](https://doi.org/10.1044/1092-4388(2005)021)
- Rau, D. V., Chang, H. H. A., & Tarone, E. E. (2009). Think or sink: Chinese learners' acquisition of the English voiceless interdental fricative. *Language Learning*, 59(3), 581–621. <https://doi.org/10.1111/j.1467-9922.2009.00518.x>
- Schmidt, A. M., & Meyers, K. A. (1995). Traditional and phonological treatment for teaching English fricatives and affricates to Koreans. *Journal of Speech, Language, and Hearing Research*, 38(4), 828–838. <https://doi.org/10.1044/jshr.3804.828>
- Schoonmaker-Gates, E. (2015). On voice-onset time as a cue to foreign accent in Spanish: Native and nonnative perceptions. *Hispania*, 98, 779–791. <https://doi.org/10.1353/hpn.2015.0110>
- Schuhmann, K. S., & Huffman, M. K. (2019). Development of L2 Spanish VOT before and after a brief pronunciation training session. *Journal of Second Language Pronunciation*, 5(3), 402–434. <https://doi.org/10.1075/jslp.18018.sch>
- Shadle, C. H. (1990). Articulatory-acoustic relationships in fricative consonants. In W. J. Hardcastle, A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 187–209). Dordrecht: Springer. [https://doi.org/10.1007/978-94-009-2037-8\\_8](https://doi.org/10.1007/978-94-009-2037-8_8)
- Shadle, C. H. (2012). Acoustics and aerodynamics of fricatives. In Cohn, A. C., Fougheron, C., Huffman, M. K., & Renwick, M. E. L. (Eds.), *The Oxford Handbook of Laboratory Phonology* (pp. 511–526). Oxford University Press.
- Shadle, C. H., & Mair, S. J. (1996). Quantifying spectral characteristics of fricatives. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, 3, 1521–1524. <https://doi.org/10.1109/ICSLP.1996.607906>
- Shadle, C. H., Mair, S. J., & Carter, J. N. (1996). Acoustic characteristics of the front fricatives. In *1st ETRW on Speech Production Modeling*, 193–196. Retrieved from <http://www.isca-speech.org/archive>
- Shen, J. (1987). “Beijinhua hekouhu ling shengmude yuyin fenqi. (北京话合口呼零声母的语音分歧) [Variation of the initial /w/ in Beijing Mandarin].” *Zhongguo Yuwen*, 5, 352–362.
- Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, 55, 779–798. <https://doi.org/10.1044/jshd.5504.779>
- Stevens, K. N. (1971). Airflow and turbulence noise for fricative and stop consonants: Static considerations. *The Journal of the Acoustical Society of America*, 50(4B), 1180–1192. <https://doi.org/10.1121/1.1912751>
- Stevens, K. N. (1985). “Evidence for the role of acoustic boundaries in the perception of speech sounds,” In *Phonetic Linguistics*, edited by V. A. Fromkin, Academic, New York, (pp. 243–255).
- Stevens, K. N. (1998). *Acoustic Phonetics* (The MIT Press, Cambridge, MA).

Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., & Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *The Journal of the Acoustical Society of America*, 91(5), 2979–3000. <https://doi.org/10.1121/1.402933>

Strevens, P. (1960). Spectra of fricative noise in human speech. *Language and Speech*, 3(1), 32–49. <https://doi.org/10.1177/002383096000300105>

Styler, W. (2014). Spectral peak Praat script. [code] [https://github.com/stylerw/styler\\_praat\\_scripts/blob/master/peak\\_spectral\\_COG.praat](https://github.com/stylerw/styler_praat_scripts/blob/master/peak_spectral_COG.praat)

Vaughn, C., Baese-Berk, M., & Idemaru, K. (2019). Re-examining phonetic variability in native and non-native speech. *Phonetica*, 76(5), 327–358. <https://doi.org/10.1159/000487269>

Wiener, S., & Shih, Y. T. (2013). Evaluating the emergence of [v] in modern spoken Mandarin. In J.-S. Zhuo (Ed.), *Toward Increased Empiricism: Studies in Chinese Linguistics* (pp. 171–187). Philadelphia, PA: John Benjamins. <https://doi.org/10.1075/scl.2.08wie>

Xie, X., & Myers, E. B. (2017). Learning a talker or learning an accent: Acoustic similarity constrains generalization of foreign accent adaptation to new talkers. *Journal of Memory and Language*, 97, 30–46. <https://doi.org/10.1016/j.jml.2017.07.005>

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. *Journal of Experimental Psychology: Human Perception and Performance*, 43(1), 206–217. <https://doi.org/10.1037/xhp0000285>

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

Zhang, Y., & Xiao, J. (2014). An analysis of Chinese students’ perception and production of paired English fricatives: From an ELF perspective. *Journal of Pan-Pacific Association of Applied Linguistics*, 18(1), 171–192. Retrieved from [eric.ed.gov/?id=EJ1047452](http://eric.ed.gov/?id=EJ1047452)

Zheng, Y., & Samuel, A. G. (2017). Does seeing an Asian face make speech sound more accented? *Attention, Perception & Psychophysics*, 79(6), 1841–1859. <https://doi.org/10.3758/s13414-017-1329-2>

Appendix

1. Production task stimuli

| Target minimal pairs |        | Filler minimal pairs |      |
|----------------------|--------|----------------------|------|
| fail                 | veil   | bark                 | park |
| fair                 | there  | base                 | pace |
| fat                  | vat    | bath                 | path |
| fault                | vault  | bay                  | pay  |
| fay                  | they   | bet                  | pet  |
| feces                | theses | big                  | pig  |
| fees                 | these  | bill                 | pill |

| Target minimal pairs |         | Filler minimal pairs |       |
|----------------------|---------|----------------------|-------|
| fin/fen              | then    | bye                  | pie   |
| fence                | thence  | came                 | game  |
| ferry                | very    | cap                  | gap   |
| few                  | view    | card                 | guard |
| file                 | vile    | cave                 | gave  |
| fine                 | thine   | coast                | ghost |
| first                | thirst  | come                 | gum   |
| focal                | vocal   | could                | good  |
| foe                  | throw   | curl                 | girl  |
| fought               | thought | deck                 | tech  |
| foul                 | vowel   | dime                 | time  |
| free                 | three   | dip                  | tip   |
| fresh                | thresh  | do                   | two   |
| fret                 | threat  | door                 | tore  |
| frill                | thrill  | down                 | town  |
| fro                  | though  | drain                | train |
| fuss                 | thus    | dry                  | try   |

Critical targets used for acoustic analysis and L1 rating

|         |        |
|---------|--------|
| Ferry   | /fɛri/ |
| Very    | /vɛri/ |
| Fail    | /feɪl/ |
| Veil    | /veɪl/ |
| Few     | /fju/  |
| View    | /vju/  |
| Fees    | /fɪz/  |
| These   | /ðɪz/  |
| Fair    | /fɛr/  |
| There   | /ðɛr/  |
| Fin/fen | /fɪn/  |
| Then    | /ðɪn/  |

|         |         |
|---------|---------|
| First   | /fə-st/ |
| Thirst  | /θə-st/ |
| Fought  | /fɔt/   |
| Thought | /θɔt/   |

2. Output of statistical models (note: All reported p-values adjusted with Bonferroni correction)

|   | $\beta$ estimate | SE   | <i>t</i> | <i>p</i> |
|---|------------------|------|----------|----------|
| Normalized amplitude:   |                  |      |          |          |
| (Intercept)   | -17.69           | 1.58 | -11.17   | <.001    |
| Voiceless   | 5.34             | 2.38 | 2.24     | .24      |
| Labiodental   | -2.15            | 2.13 | -1.01    | .99      |
| Voiceless*Labiodental   | -5.33            | 2.97 | -1.80    | .54      |
| lmer(norm.amp~voicing*articulation+(1 speaker)+(1 item),data) |                  |      |          |          |
| Relative amplitude:   |                  |      |          |          |
| (Intercept)   | -17.03           | 1.29 | -13.39   | <.001    |
| Voiceless   | -1.25            | 1.69 | -0.74    | .99      |
| Labiodental   | -2.11            | 1.52 | 1.39     | .99      |
| Voiceless*Labiodental   | 4.27             | 2.11 | 2.03     | .36      |
| lmer(rel.amp~voicing*articulation+(1 speaker)+(1 item),data)  |                  |      |          |          |
| Normalized duration:  |                  |      |          |          |
| (Intercept)   | 0.13             | 0.02 | 7.77     | <.001    |
| Voiceless   | 0.07             | 0.02 | 3.52     | .03      |
| Labiodental   | 0.02             | 0.01 | 1.24     | .99      |
| Voiceless*Labiodental   | -0.04            | 0.02 | -1.66    | .72      |
| lmer(norm.dur~voicing*articulation+(1 speaker)+(1 item),data) |                  |      |          |          |
| Fricative duration:   |                  |      |          |          |
| (Intercept)   | 68.79            | 9.69 | 7.09     | <.001    |
| Voiceless   | 49.75            | 7.85 | 6.33     | <.001    |
| Labiodental   | 12.33            | 7.04 | 1.75     | .81      |
| Voiceless*Labiodental   | -30.98           | 9.79 | -3.17    | .06      |
| lmer(fric.dur~voicing*articulation+(1 speaker)+(1 item),data) |                  |      |          |          |



|   | $\beta$ estimate | SE     | <i>t</i> | <i>p</i> |
|---|------------------|--------|----------|----------|
| Spectral peak:  |                  |        |          |          |
| (Intercept)   | 5830.01          | 92.72  | 62.88    | <.001    |
| Voiceless   | 1956.41          | 115.20 | 16.98    | <.001    |
| Labiodental   | 140.10           | 103.54 | 1.35     | .99      |
| Voiceless*Labiodental   | 143.76           | 143.73 | 1.00     | .99      |
| lmer(peak~articulation*voicing+(1 speaker)+(1 item),data)     |                  |        |          |          |
| Spectral centroid:  |                  |        |          |          |
| (Intercept)   | 4638.24          | 128.41 | 36.12    | <.001    |
| Voiceless   | 1178.53          | 187.89 | 6.27     | <.001    |
| Labiodental   | -851.42          | 168.47 | 1.62     | .99      |
| Voiceless*Labiodental   | 497.83           | 234.18 | 2.13     | .30      |
| lmer(centroid~voicing*articulation+(1 speaker)+(1 item),data) |                  |        |          |          |


Releveled logistic regression models on correct fricative identification (phoneme talker block models)

|  | $\beta$ estimate | SE   | <i>Z</i> | <i>p</i> |
|--|------------------|------|----------|----------|
| (Intercept: /v/ reference level)   | 0.54             | 0.13 | 4.33     | <.001    |
| /v/ – /f/  | 0.27             | 0.13 | 2.05     | .04      |
| /v/ – /θ/  | 0.49             | 0.18 | 2.64     | .008     |
| /v/ – /ð/  | 0.51             | 0.16 | 3.10     | .001     |
| Accent Rating  | -0.08            | 0.11 | -0.73    | .45      |
| Accent Rating: /v/ – /f/   | 0.07             | 0.13 | 0.57     | .56      |
| Accent Rating: /v/ – /θ/   | -0.30            | 0.18 | -1.06    | .10      |
| Accent Rating: /v/ – /ð/   | 0.27             | 0.16 | 1.63     | .09      |
| glmer(onset.accuracy~phoneme*rating+(1 rater)+(1 item), family="binomial", optimizer="bobyqa") |                  |      |          |          |
| (Intercept: /θ/ – reference level)   | 1.03             | 0.16 | 6.39     | <.001    |
| /θ/ – /f/  | -0.22            | 0.16 | -1.35    | .17      |
| /θ/ – /v/  | -0.49            | 0.18 | -2.63    | .001     |
| /θ/ – /ð/  | 0.01             | 0.19 | 0.09     | .92      |
| Accent Rating  | -0.38            | 0.15 | -2.48    | .01      |

|   | $\beta$ estimate | SE   | Z     | p     |
|---|------------------|------|-------|-------|
| Accent Rating: /θ/ – /f/  | 0.37             | 0.16 | 2.25  | .02   |
| Accent Rating: /θ/ – /v/  | 0.30             | 0.18 | 1.60  | .10   |
| Accent Rating: /θ/ – /ð/  | 0.57             | 0.19 | 2.95  | .003  |
| glmer(onset.accuracy~phoneme*rating+(1 rater)+(1 item), family="binomial",<br>optimizer="bobyqa") |                  |      |       |       |
| (Intercept: /ð/ – reference level)  | 1.05             | 0.13 | 7.74  | <.001 |
| /ð/ – /f/   | 0.24             | 0.14 | –1.72 | .08   |
| /ð/ – /v/   | –0.50            | 0.16 | –3.10 | .001  |
| /ð/ – /θ/   | –0.01            | 0.19 | –0.09 | .92   |
| Accent Rating   | 0.19             | 0.12 | 1.53  | .12   |
| Accent Rating: /ð/ – /f/  | –0.19            | 0.14 | –1.42 | .15   |
| Accent Rating: /ð/ – /v/  | 0.27             | 0.16 | –1.67 | .09   |
| Accent Rating: /ð/ – /θ/  | –0.57            | 0.19 | –2.96 | .003  |
| glmer(onset.accuracy~phoneme*rating+(1 rater)+(1 item), family="binomial",<br>optimizer="bobyqa") |                  |      |       |       |

Address for correspondence

Seth Wiener  
Department of Modern Languages  
Carnegie Mellon University  
5000 Forbes Ave.  
Pittsburgh, PA, 15213  
United States  
sethw1@cmu.edu

 <https://orcid.org/0000-0002-7383-3682>

## Co-author information

Zhe Gao  
Department of Modern Languages  
Carnegie Mellon University  
zheg@andrew.cmu.edu

Xiaomeng Li  
Department of Modern Languages  
Carnegie Mellon University  
xiaomenl@andrew.cmu.edu

Zhiyi Wu  
Department of Modern Languages  
Carnegie Mellon University  
wuzhiyi.jenny@outlook.com

## Publication history

Date received: 1 December 2020

Date accepted: 20 January 2022

Published online: 13 May 2022