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

Yue Zheng, Paul Iverson
University College London
yue.zheng.15@ucl.ac.uk, p.iverson@ucl.ac.uk

ABSTRACT

This study examined how transition cues affect the perception and categorization of /f/ and /s/ by native Cantonese, Mandarin, and English speakers. Previous work had suggested that speakers of languages with small fricative inventories are much less dependent on formant transitions, likely because they can distinguish among their native-language fricatives using spectra alone. The present study investigated this issue by combining behavioural and EEG measures (a phoneme monitoring task and P300 measure integrated in an active oddball paradigm) for language groups that differed in their fricative inventories. The stimuli were spliced with vowels such that the formant transitions were congruent or incongruent with the fricatives. The results revealed that all groups attended to formant transitions when identifying fricatives, despite their differing language backgrounds and fricative inventories.

Keywords: fricative formant transitions, cross-language difference, coarticulatory cues, L2 perception

1. INTRODUCTION

Although fricative categorisation is thought to be primarily driven by aspects of the fricative spectra (e.g., spectral moments), it can also be affected by formant transitions in the surrounding vowels [1, 2, 3, 4, 5, 6]. Formant transition cues are likely not equally important for all the fricatives; some fricatives are more spectrally distinct (e.g., English /s/) so that they can be distinguished from other fricatives (e.g., English /f/ and /θ/) based on frication alone [1, 2, 3, 4]. The use of formant cues can also be language-specific [5]; speakers of languages with small fricative inventories and without spectrally similar fricatives (e.g., Dutch and German) are less sensitive to formant transitions, and speakers of languages with more spectrally similar fricatives rely more heavily on transitional cues (e.g., Spanish and Polish). In general, formant transitions appear to be a secondary cue for fricative identification, given that they are exploited mostly when spectral cues are not sufficient.

Native-language cue weightings and categories can continue to be influential when perceiving unfamiliar fricatives rather than listeners being able to flexibly adjust their ways of cue processing [5, 6]. For example, Cantonese and Mandarin native speakers both have difficulty perceiving and producing the English /θ/. However, these two language groups assimilate English /θ/ to different categories (i.e., Cantonese /f/ and Mandarin /s/), which is surprising because both languages have /f/ and /s/ [7, 8, 9, 10, 11]. This could occur because /f/ and /s/ fricatives have slightly different spectra in Cantonese and Mandarin (e.g., English /θ/ may be more like Cantonese /f/) [11]. However, we hypothesised that this difference could be explained by formant transitions; English /f/ and /θ/ are spectrally similar, but /θ/ involves a tongue articulation that may make its transitions more like /s/. Cantonese speakers may not use formant transitions because of their small fricative inventory (as shown in Table 1), but Mandarin speakers have an additional fricative contrast /s/-/ʂ/ that could make them more reliant on transitions and thus hear English /θ/ as /s/.

Table 1. A comparison of English, Mandarin and Cantonese inventories of voiceless fricatives (excluding the glottal fricative /h/ from the English and Cantonese inventories, as it does not involve a constriction within the oral cavity, and is considered different from the other fricatives [9, 12]).

ENGLISH	f	θ	s	ʃ			
MANDARIN	f		s		ʂ	ɕ	x
CANTONESE	f		s				

The present study investigated possible cross-language differences in the weighing of formant transition cues during fricative perception by Mandarin, Cantonese and English speakers, in order to understand the assimilation patterns of /θ/. We used an active oddball paradigm, in which a series of stimuli are each identified as “non-target” or a relatively infrequent “target” (i.e., /s/ or /f/). The stimuli were either cross spliced (i.e., frication spliced into a vowel context from another fricative, such as /f/ replacing the frication in /sɑ/) or identity spliced (i.e., /f/ replacing the frication of a /fɑ/ syllable, such that the splicing operation was the same for the two types of stimuli). The identification of targets was measured both behaviourally (i.e., identification and reaction times) and in terms of the P300 ERP from

EEG recordings. P300 shows reliability in studies of phonological processing. It appears to be sensitive to the phonological categories during phonological processing, and it preserves some effects of acoustic details [13, 14, 15, 16]. Using P300 to investigate perceptual strategy and cross-language differences in fricative perception should be able to provide evidence of how language experience affects phonological processes, and to support the behavioural results.

2. METHOD

2.1. Participants

This experiment recruited 12 native Southern British English speakers, 12 native Northern Mandarin Chinese speakers, and 12 native Hong Kong Cantonese speakers, who were all right-handed adults between 18 to 30 years old. They reported no history of hearing, learning, or language impairment, and no history of neurological disorders. The native English speakers had no knowledge of either Mandarin or Cantonese. The Mandarin and Cantonese speakers started learning English after 5 years old, and had been exposed to an English-speaking environment for less than 2 years.

2.2. Stimuli

Four female native speakers of each target language produced the stimuli in their native language for this experiment. They read out CV syllables naturally (C were fricatives, and V were /a ɑ/). The length of each stimulus was equated to 0.55 s (with 0.15 s of frication) using Praat. The recorded syllables with the target fricatives were then spliced in two ways, identity spliced and cross spliced. An identity-spliced stimulus had its fricative replaced by the same fricative of another token from the same language. A cross-spliced fricative was replaced by the other target fricative of a token from the same language (e.g. the /f/ of an English /fɑ/ was replaced by a /s/ from an English /sɑ/). The point of splicing for the recorded target fricatives was at the zero-crossing point at the end of frication and the beginning of harmonic structure, following the same method adopted by Wagner et al. [5].

2.3. Procedure

ERPs were recorded using a 64-channel BioSemi ActiveTwo system with 2048 Hz sampling rate. Participants were informed what the target phoneme was before the start of each block. They were asked to press the target button (marked with a sticker) on a button box as soon as they could identify a target stimulus, and to press the other button for any other stimuli. They were reminded that the next sound would play shortly after they pressed a button. The experiment was divided into six blocks, three with /f/ and three with /s/. A filler condition (i.e., other non-target fricatives) was added to maximize the amplitude of P300, as the target-to-target interval needed to be longer than 6 s, with 2 to 4 fillers in between. The probability of a target stimulus in each block was 25%. However, it was assumed that the cross-spliced targets would sometimes not be perceived as a target, so that the target probability for the participants should vary between 12.5% and 25%.

Table 2: The stimuli of the experiment.

Target	Stimulus Type		
	Identity -spliced (ID)	Cross-spliced (CR)	Filler
/f/	/f/-ID	/f/-CR	Mandarin /ʂ/, English /ʃ/, voiced fricatives
/s/	/s/-ID	/s/-CR	Mandarin /ʂ/, English /ʃ/ and /θ/, voiced fricatives

2.4. Data analysis

Behavioural performance was analysed as the percentage of targets identified, and the median reaction time between a button press and a stimulus onset under each condition. The P300 response analysed was the mean amplitude between 0.3 s and 0.8 s after stimulus onset, recorded at all the parietal and mid-line channels. Mixed-model analyses were used with *subject language* as the between-subject variable, and with *target* and *stimulus type* as within-subject variables.

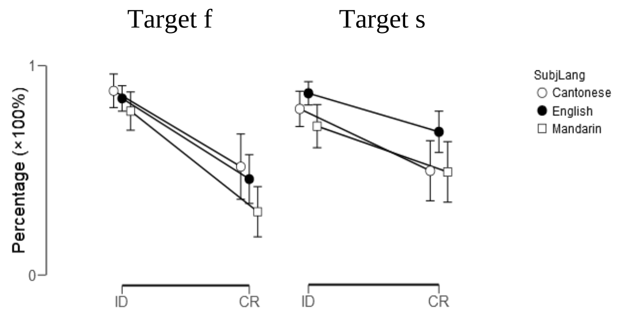
3. RESULTS

3.1. Behavioural results

3.1.1. Percentage of correct target identification

The identification percentages indicated that all language groups were sensitive to cross splicing, with fewer target identifications for the cross-spliced stimuli. There was a significant main effect of stimulus type, $F(1,33)=215.33$, $p<0.001$, with more target identifications for identity-spliced than cross-spliced stimuli. There was a main effect of subject language, $F(2,33)=4.44$, $p<0.05$, with English speakers being slightly more frequent at identifying targets. There was a significant interaction between target and stimulus type, $F(1,33)=22.60$, $p<0.001$; /f/ was more affected by cross splicing than /s/. However, there were no significant interaction involving subject language, $p > 0.05$, indicating that the effect of transitions was similar between the language groups.

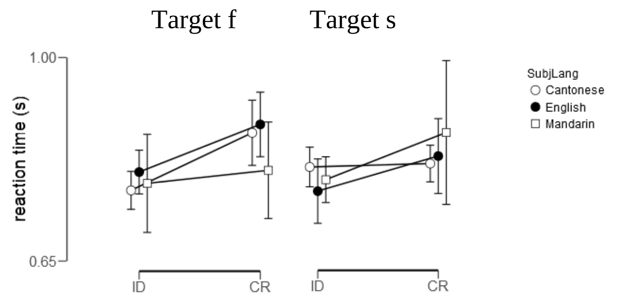
Figure 2: Average percentages of correct target identification for participants of three language groups under different conditions.



3.1.2. Reaction time

The only significant main effect was stimulus type, $F(1,33)=15.52$, $p<0.001$. That is, listeners were slower at identifying the target for cross-spliced stimuli, even when considering only trials in which a target was identified. All the interactions involving subject language were non-significant, $p > 0.05$.

Figure 3: Average reaction time of participants of three language groups under different conditions.



3.2. P300 results

The only significant main effect was stimulus type, $F(1,33)=7.69$, $p<0.01$, with greater P300 responses for identity-spliced stimuli. All the interactions involving subject language were non-significant, $p > 0.05$.

Figure 4: Average P300 amplitudes of three groups of participants under different conditions.

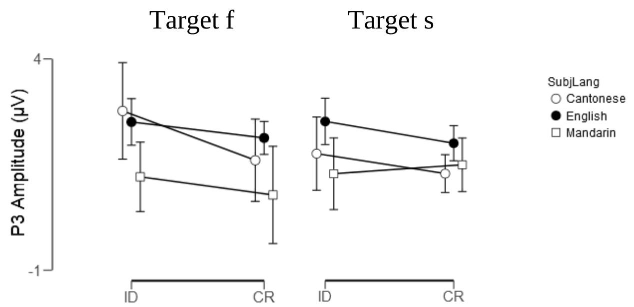
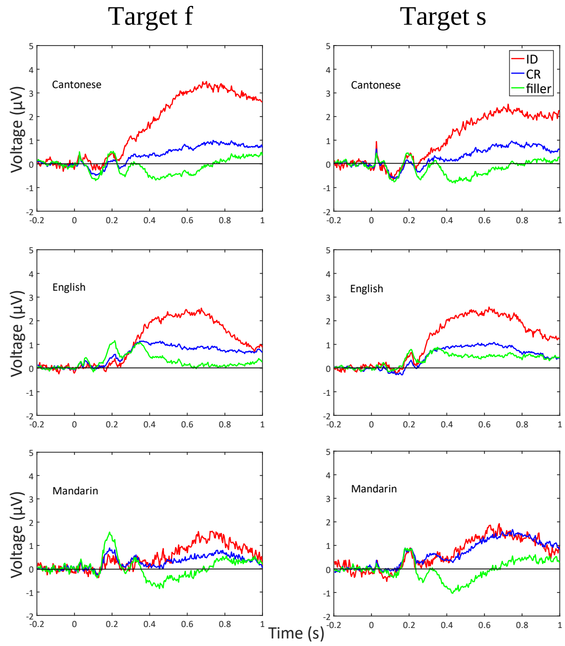


Figure 5. Grand-average even-related potential (ERP) waveforms, averaged across parietal and mid-line electrodes as a function of stimulus type and subject language.



4. DISCUSSION

The significant main effects of stimulus type in both behavioural and EEG measures indicates that Cantonese, Mandarin, and English speakers are all affected by mismatching formant transitions, demonstrated by lower accuracy rates, longer reaction time, and a smaller mean P300 amplitude under the cross-spliced condition.

The finding that Mandarin speakers make use of formant transition cues supports our original hypothesis; we expected that as Mandarin speakers would use formant transitions in fricative identification because they have more apical fricatives. However, it was unexpected to find that the Cantonese-speaking participants also attended to the formant transitions, despite their smaller fricative inventory. Like Dutch, Cantonese has only two fricative categories in front of the alveolar position: /f/ and /s/ [5, 9]. However, Dutch and Cantonese are from different language families with different ways of processing segmental and coarticulatory cues. Cantonese and Mandarin are tonal languages, in which tone segments are carried by syllables, in which case the sense of unity of syllables may be reinforced by the tones. Several studies on Mandarin have claimed that syllables are stored and retrieved as a whole unit, and syllables, instead of phonemic segments, play a primary role in speech processing and production [17, 18, 19]. Cantonese is presumably the same [19]. In contrast, Indo-European languages (e.g., Dutch), treat phonemic segments as units, and treat syllables as compositions built up with the units.

This difference could lead to their different perception of formant transitions.

In this study there was a possible effect of English learning experience that was difficult to eliminate, which may have increased the necessity of using formant transitions. Currently, no other studies have provided sufficient evidence to determine how much training is needed to learn to make more use of formant transitions during second language learning.

The P300 results demonstrate that the manipulated acoustic information, i.e. the mismatched formant transitions, is preserved at a post-perceptual level, which is comparable to previous P300 results [13]. P300 amplitude is considered to vary with the amount of attentional resource allocated in the task [23]; the lower amplitude under cross-spliced conditions across language groups may indicate that their attention was divided after a cross-spliced stimulus, as the listeners may have continued to search for additional information elsewhere to make up for the difficult cues from the stimuli.

In conclusion, this study found that the Cantonese, Mandarin, and English native speakers were affected by mismatching formant transition cues when identifying fricatives, regardless of their native fricative inventories. This suggests that the motivation to use formant cues is more complicated than the content or the size of a fricative inventory.

5. REFERENCES

[1] Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and speech, 1*(1), 1-7.

[2] Strevens, P. (1960). Spectra of fricative noise in human speech. *Language and speech, 3*(1), 32-49.

[3] Nittrouer, S., & Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech, Language, and Hearing Research, 30*(3), 319-329.

[4] 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.

[5] Wagner, A., Ernestus, M., and Cutler, A. (2006). Formant transitions in fricative identification: The role of native fricative inventory. *The Journal of the Acoustical Society of America, 120*(4), 2267-2277.

[6] Scharenborg, O., Weber, A., & Janse, E. (2015). Age and hearing loss and the use of acoustic cues in fricative categorization. *The Journal of the Acoustical Society of America, 138*(3), 1408-17.

[7] Jiang, Y. (1995). Chinglish and China English. *English Today, 41*(11), 51-53.

[8] Hung, T. T. N. (2000). Towards a phonology of Hong Kong English. *World Englishes, 19*, 337–356.

[9] Chan, A. Y. & Li, D. C. (2000). English and Cantonese phonology in contrast: Explaining Cantonese ESL learners' English pronunciation problems. *Language Culture and Curriculum, 13*(1), 67-85.

[10] Peng, L. & Setter, J. (2000). The emergence of systematicity in the English pronunciations of two Cantonese speaking adults in Hong Kong. *English World-Wide, 21*(1), 81-108.

[11] Zheng, Y. & Iverson, P. (2016). *The assimilation of the English /θ/ and how it varies cross-linguistically*. Unpublished manuscript.

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

[13] Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological science, 21*(10), 1532-1540.

[14] Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological psychology, 41*(2), 103-146.

[15] Newman, R. L., Connolly, J. F., Service, E., & Mcivor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. *Psychophysiology, 40*(4), 640-647.

[16] Fosker, T., & Thierry, G. (2004). P300 investigation of phoneme change detection in dyslexic adults. *Neuroscience letters, 357*(3), 171-174.

[17] Chen, J. Y. (1999). The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied psycholinguistics, 20*(2), 289-301.

[18] Chen, J. Y. (2000). Syllable errors from naturalistic slips of the tongue in Mandarin Chinese. *Psychologia: An International Journal of Psychology in the Orient, 43*, 15–26.

[19] Chen, J. Y., O'séaghdha, P. G., & Chen, T. M. (2016). The primacy of abstract syllables in Chinese word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(5), 825.

[20] Xu, Y., & Wang, M. (2009). Organizing syllables into groups—Evidence from F0 and duration patterns in Mandarin. *Journal of phonetics, 37*(4), 502-520.

[21] Mcmurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*(2), B33-42.

[22] Mcmurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language, 60*(1), 65-91.

[23] Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology, 118*(10), 2128-2148.