

BRIEF RESEARCH REPORT

Temporal features of word-initial /s/+stop clusters in bilingual Mandarin–English children and monolingual English children and adults*

JING YANG

Department of Communication Sciences and Disorders, University of Central Arkansas

(Received 5 August 2016 – Revised 30 December 2016 – Accepted 12 April 2017 – First published online 18 May 2017)

ABSTRACT

This study investigated the durational features of English word-initial /s/+stop clusters produced by bilingual Mandarin (L1)–English (L2) children and monolingual English children and adults. The participants included two groups of five- to six-year-old bilingual children: low proficiency in the L2 (Bi-low) and high proficiency in the L2 (Bi-high), one group of age-matched English children, and one group of English adults. Each participant produced a list of English words containing /sp, st, sk/ at the word-initial position followed by /a, i, u/, respectively. The absolute durations of the clusters and cluster elements and the durational proportions of elements to the overall cluster were measured. The results revealed that Bi-high children behaved similarly to the English monolinguals whereas Bi-low children used a different strategy of temporal organization to coordinate the cluster components in comparison to the English monolinguals and Bi-high children. The influence of language experience and continuing development of temporal features in children were discussed.

INTRODUCTION

A consonant cluster refers to a sequence of adjacent consonants that can appear at the onset or coda position in a syllable. This type of phonetic structure is

[*] This research was supported, in part, by the Graduate School's Alumni Grants for Research and Scholarship fund at The Ohio State University. The author would like to thank Dr Robert Fox for the constructive comments during the preparation of the manuscript. Address for Correspondence: Jing Yang, Department of Communication Sciences and Disorders, University of Central Arkansas, 201 Donaghey Ave, Conway AR 72035. tel: 501-450-5486; e-mail: [jyang@uca.edu](mailto: jyang@uca.edu)

widely used in many of the world's languages (Greenberg, 1965; Locke, 1983). English has a large pool of consonant clusters that vary in length and complexity (Hultzén, 1965). Early studies on the development of consonant clusters in English monolingual children have shown that consonant clusters emerge as early as two years of age and that the majority of clusters are mastered before seven years of age (e.g. Kirk & Demuth, 2005; McLeod, Van Doorn & Reed, 2001a, 2001b; Smit, Hand, Freilinger, Bernthal & Bird, 1990). More recently, a number of phonological studies examined the acquisition of English consonant clusters by adult L2 learners from various language backgrounds (e.g. Carlisle, 1988; Eckman & Iverson, 1993; Hansen, 2001; Jayaraman, 2010; Nguyen, 2008). However, little research has been done to explore the acoustic-phonetic features of consonant clusters produced by English-learning children. The present study, then, attempts to extend previous studies by investigating the temporal features of English word-initial /s/+stop consonant clusters produced by bilingual Mandarin-English children and monolingual English children and adults.

Among the English consonant clusters, word-initial /s/+stop clusters are of particular interest to the present study. Over the past few decades, much research attention has been paid to the acquisition and production of word-initial clusters. Previous studies reported that while word-initial singletons were acquired earlier than word-final singletons, word-initial clusters were generally produced with lower accuracy and were acquired later than word-final clusters in young children (e.g. Kirk & Demuth, 2005; Levelt, Schiller & Levelt, 2000). The /s/+stop cluster is a two-element phonetic structure composed of a fricative and a stop sound. These two types of consonants represent two main categories of manners of articulation in most of the world's languages. However, word-initial /s/+stop clusters are phonologically 'atypical' in a number of ways (Treiman, 1986; Yavaş & Someillan, 2005). It is composed of two obstruents that are normally disallowed at syllable onsets. The stops in /s/+stop clusters are not the typical voiced stops /b, d, g/ or the typical aspirated voiceless stops /p, t, k/ as shown in singletons. The components in /s/+stop clusters are more closely adhered to each other than the components in other clusters (Fudge, 1969; Kohler, 1968). Phonetically, the combination of these two types of consonants in a cluster shows distinct phonetic features from the singletons of each type of sounds (Klatt, 1975; Menyuk & Klatt, 1975; Schwartz, 1970).

A consonant cluster is not simply a combination of two or more single consonants. Producing a consonant cluster requires complex coordination of articulatory gestures for two or more components. Borden and Gay (1979) compared the temporal organization of articulatory movement of consonants in /s/+stop clusters. They found that the tongue was not moving as a whole. Instead, the anterior and posterior portion acted as

multiple articulators and moved separately for different phonemes. The temporal aspect represents one of the most basic acoustic properties in speech production. It also functions as an important indicator of the development of articulatory coordination and speech timing control (Gilbert & Purves, 1977). The durational features of consonant clusters and their components are affected by multiple factors. The components' phonetic features of place, manner, and voicing, the number of components in a cluster, and the position of one component relative to the other are all related to the intrinsic properties of the cluster components and the internal structure of the cluster (Byrd, 1996; Klatt, 1975; O'Shaughnessy, 1974). Previous studies have shown that the syllable structure and contextual environment also play a role in conditioning the durations of the components in a cluster (Byrd, 1996; Umeda, 1977).

In addition to the above recognized factors, a speaker's language experience may also have an impact on the temporal features of a consonant cluster. Although there is a lack of direct evidence showing the correlation between the temporal features of English consonant clusters and the L2 experience, substantial work has reported that adult L2 speakers generally produced speech segments (vowels and single consonants), syllables, words, and utterances in English with longer duration and slower speaking rate than native English speakers (e.g. Baker, Baese-Berk, Bonnasse-Gahot, Kim, Van Engen & Bradlow, 2011; Flege, 1991; Guion, Flege, Liu & Yeni-Komshian, 2000; Riegenbach, 1991). The longer durations of the speech segments and utterances contributed to the degree of foreign accent in non-native speakers. Researchers assumed that the longer duration in speech of L2 speakers indicated that non-native speakers may need longer time to process the L2 and require more processing resources to reduce the interference of the L1. Further examination revealed that the large amount of variation in the temporal features among L2 speakers may be associated with the age of learning (Guion *et al.*, 2000; Trofimovich & Baker, 2006) and the level of language proficiency (Riazantseva, 2001). The general trend is that L2 speakers who started to learn a new language at a younger age and have a higher proficiency in the L2 are more likely to produce native-like temporal features and can better adjust their speech behavior to approximate native speakers of the L2.

THE CURRENT STUDY

The alveolar fricative /s/ and stops /p, t, k/ are commonly used in both Mandarin and English. However, Mandarin lacks the structure of the consonant cluster. We are interested in how native Mandarin children form the phonetic structure in which all components are familiar but the

structure itself does not occur in their native language. To address this question, the absolute duration of word-initial /s/+stop clusters and cluster components, as well as the durational proportion of each component to the entire cluster, were compared among bilingual Mandarin–English children and monolingual English children and adults. The absolute duration describes the amount of time used for gestural movement of individual elements in individual speakers. This measure provides information on the amount of temporal articulatory constraint a speaker encounters in producing complex phonetic structures. The durational proportion defines the weight of individual elements to the entire cluster. This measure provides a normalized measurement to compare the weight of individual components and the relationship of one component to the other across different speakers. More importantly, the relative duration of proportion of each element to the entire cluster enables us to examine the mechanism of temporal organization in the production of consonant clusters.

The present study included two groups of sequential bilingual Mandarin–English children in the same age range. One group of children (Bi-low) had a short period of residency in the US and had low proficiency in English (L2). The other group of children (Bi-high) had mostly lived in the US from birth, but they started to learn English after the acquisition of Mandarin. This group of children had a high proficiency in English. Both English monolingual children (EC) and adults (EA) were recruited as control groups. Previous phonetic studies revealed longer durations of speech units in adult L2 learners and more native-like speech performance in early L2 learners (Baker *et al.*, 2011; Guion *et al.*, 2000; Riazantseva, 2001; Trofimovich & Baker, 2006). According to these findings, when comparing the durational features of consonant clusters in bilingual children with those of English monolingual children, we tentatively predict that the Bi-high children may approximate English monolingual children in the durational features of consonant clusters. In particular, they may show similar absolute durations and durational proportions to English monolingual children. In contrast, the Bi-low children are likely to produce English consonant clusters with longer durations than English monolingual children. However, since there is little research on the temporal organization of English clusters in young English learners at the beginning stage of L2 learning, how the Bi-low children assign temporal features to each element is of particular interest and to be examined in the present study.

When comparing the bilingual children with English-speaking monolingual adults, the differences between children and adults in the timing control of cluster production should be noted. Tingley and Allen (1975) suggested that there may be a common timing control mechanism in children at different ages, but that the temporal variability tended to

decrease as a function of age. McLeod and colleagues (2001b) found that the consonant clusters produced by young children may demonstrate distinct features from those in adult speakers. Some early phonetic studies revealed that young children produce the components in initial clusters with longer duration (Menyuk & Klatt, 1975), and children may utilize different timing strategies to organize consonant clusters compared to adults (Gilbert & Purves, 1977). Given the inferior timing control abilities in children relative to adults, we predict that the English monolingual children will produce the clusters and cluster elements with longer durations than the English monolingual adults. Meanwhile, the durational proportions of individual elements to the clusters in the English monolingual children may differ from those in the adults. Considering the potential differences between children and adult speakers, when we compare the bilingual children with English monolingual adults, we predict that both Bi-low and Bi-high children will produce the clusters and cluster elements with longer durations than English monolingual adults. But how the Bi-low and Bi-high children temporally organize the cluster components in comparison to English monolingual adults is not clear and will be examined.

METHODS

Speakers

Twenty-four children and six adults participated in the present study. The twenty-four children included fifteen sequential Mandarin–English bilingual children at five to six years of age, and nine age-matched English monolingual children. The bilingual Mandarin–English children were divided into two groups based on their proficiency in English (see Table 1 for detailed information). The Bi-low group included seven children born and raised in China who had lived in the US (central Ohio region) for less than six months. Some Bi-low children started to learn English when they went to kindergarten in China, but their exposure to and usage of English were limited. The Bi-high group included eight children born and raised in the US (central Ohio region). These Bi-high children were raised in a near-monolingual environment and learned Mandarin as the first language from family contact and interaction with individuals in the local Mandarin-speaking community. They had very limited exposure to English before three years of age, when most of them went to English daycare or kindergarten. By the time of recording, these children had gained intensive experience with English for about three years. For both groups of the bilingual children, their parents came from northern dialect regions of China, and Mandarin was the daily life language used in their families. The nine English monolingual children were born and raised in

TABLE 1. *Characteristics of subgroups of bilingual Mandarin–English children: subject group, subject number, gender, age, length of residency in the US (LOR_in_US), age of learning in English (AOL_E), English usage, age of learning in Mandarin (AOL_M), Mandarin usage*

Group	SubNum	Gender	Age	LOR_in_US (in months)	AOL_E	E_usage	AOL_M	M_usage
Bi-high	1	F	5;11	58	3;6	50%	since birth	50%
	2	F	5;7	67	4;0	50%	since birth	50%
	3	M	5;5	65	3;7	80%	since birth	20%
	4	F	5;10	70	2;1	70%	since birth	30%
	5	F	5;10	47	3;0	70%	since birth	30%
	6	F	5;6	66	2;6	70%	since birth	30%
	7	M	5;1	45	2;0	70%	since birth	30%
	8	F	5;7	66	3;0	70%	since birth	30%
Bi-low	1	F	5;10	5	4;6 (in China), 5;5 (in US)	40%	since birth	60%
	2	M	5;8	3	4 (in China); 5;5 (in US)	40%	since birth	60%
	3	M	5;8	3	5;5 (in US)	30%	since birth	70%
	4	M	5;2	1;5	4;0 (in China); 5;0 (in US)	30%	since birth	70%
	5	F	5;6	6	3;0 (in China), 5;0 (in US)	20%	since birth	80%
	6	F	6;3	5	5;11 (in US)	20%	since birth	80%
	7	F	5;10	4	5;6 (in US)	40%	since birth	60%

the central Ohio region and had no or very little exposure to other English dialects. Likewise, the six adults were also from the central Ohio region and they were the mothers of the English monolingual children in the present study. No speech or language disorders were reported by any of the participants or their caregivers.

Speech material

The speech material included a list of English monosyllabic or disyllabic words: *sparkler*, *speaker*, *spoon*, *stop*, *stick*, *stew*, *sky*, *ski*, *school bus*. These words contained three word-initial /s/+stop clusters each followed by three vowels /a, i, u/. Another multisyllabic word, *Scooby-doo*, that represents the combination of /sk/+/u/, was also included in the speech material, but some participants did not produce this stimulus word in an appropriate way. Those tokens were excluded from further analysis. Selection of the target words followed these rules: (i) the words were easy for children to recognize; (ii) the words could be presented in pictures; and (iii) the words frequently occurred in children's speech.

Procedure

Prior to the recording experiment, each participant filled out a questionnaire regarding their language usage and background. The questionnaires of child speakers were completed by their parents. The recording took place in a quiet room in which each speaker was seated in front of a laptop computer wearing a Shure SM10A head-mounted microphone situated approximately 1 inch from the subject's mouth. A visual-auditory word repetition task was used to collect speech samples under control of a custom Matlab program. The entirety of the recording experiment included two blocks. In each block, randomly ordered pictures representing target words were presented on the computer screen followed by an audio prompt produced by a native adult speaker. The same random order was used for both recording blocks. The participants were then asked to repeat each word immediately after the prompt. Speech samples were recorded directly onto a hard drive disk with a 16-bit quantization rate and 44.1 kHz sampling rate.

Data analysis

All speech samples of each participant were subject to acoustic analysis followed by statistical tests. Durational features were analyzed on each token of the individual participants. The durational measurements included the absolute duration of the entire cluster, the fricative /s/, stop gap, stop burst, and voice onset time (VOT). The durational proportions of the fricative /s/, stop gap, and stop burst to the entire cluster were also

calculated. The landmark locations of the segments were determined by hand on the basis of the waveform, along with a visual check of the spectrogram through Adobe Audition 1.0. The onset of the fricative was set at the point of the beginning of frication. The offset of the fricative was set at the point where no frication energy was visible on the waveform, which was also regarded as the start of the stop closure. The end of the stop closure was also the point where the release burst started. Note that some stops had multiple bursts; the offset of the release burst was set at the point where the last burst ended. The onset of voicing was set at the point where the first period of vocalization occurred. For the tokens with voiceless stops, the onset of voicing was the same point as the onset of the following vowel. Although rare, there were some tokens with a voiced stop in which the onset of voicing and the onset of vowel were separately marked.

Prior to the statistical analysis, subject means were calculated for each temporal measurement of each cluster in each vowel context. Then, three-way repeated measures ANOVAs were conducted with the speaker group as the between-subject factor and the vowel context and cluster as the within-subject factors. A post-hoc test (Bonferroni with adjustment for multiple comparisons) was used when a significant main effect was found. The differences between the four groups of speakers were of primary interest to the present study. As no group-involved interaction effect was significant, the cluster by vowel interaction effect on certain measurements is not reported and discussed here.

RESULTS

Absolute duration of consonant cluster and components

Figure 1 presents the means and standard errors of cluster durations. The English monolingual adults produced shorter durations for all three clusters than the bilingual children and English monolingual children. Averaged across vowel contexts and stop places, the mean cluster duration was 256 ms for English monolingual adults, 311 ms for Bi-low children, 329 ms for Bi-high children, and 337 ms for English monolingual children. In contrast to the Bi-low children, the Bi-high children showed a greater approximation to the English monolingual children for most clusters in different vowel environments. Moreover, the Bi-high and English monolingual children, rather than the Bi-low children, demonstrated a pattern of cluster duration across the stop places and vowel contexts more similar to the English monolingual adults. The three-way repeated measures ANOVA confirmed the group differences (see Table 2). The post-hoc comparison demonstrated that the Bi-high and English

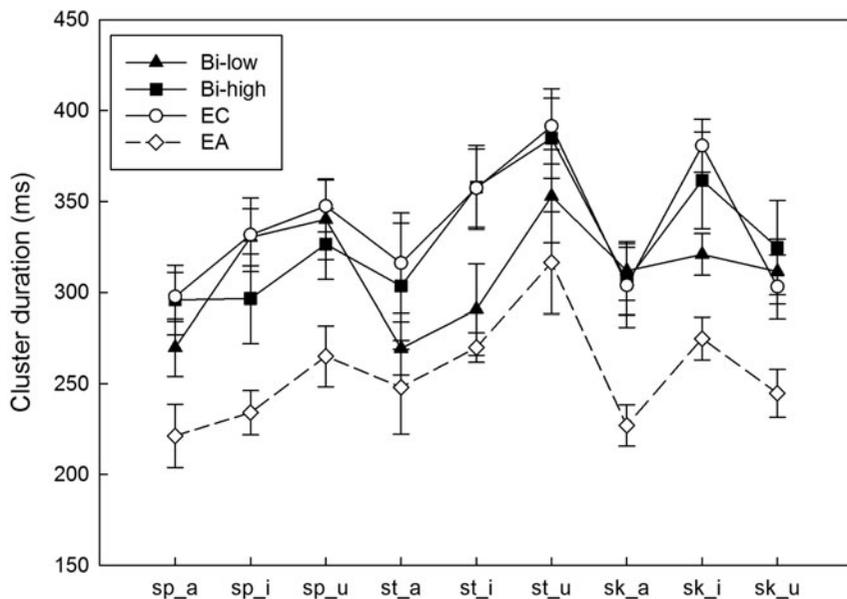


Fig. 1. The mean and standard error of cluster durations for each /s/+stop cluster followed by each of the three vowels /a, i, u/ in each group of speakers.

monolingual children produced a significantly longer cluster duration than the English monolingual adults.

The means and standard errors of the duration of fricative /s/ are presented in Figure 2. The average duration of /s/ across vowel contexts and stop places was 203 ms in Bi-high children and 195 ms in English monolingual children, which were longer than the 162 ms in Bi-low children and 168 ms in English monolingual adults. Compared to the Bi-low children, the variation of fricative duration across the clusters and vowel contexts in the Bi-high and English monolingual children was more similar to the English monolingual adults. The statistical analysis yielded significant main effects of cluster and vowel (see Table 2). However, no group difference was found.

A stop gap is the acoustic representation of the period in which the articulators move to the point of constriction and form the closure. As shown in Figure 3, the English monolingual adults produced a shorter stop gap (61 ms) than all three groups of children. The Bi-low children produced the longest stop gap (122 ms) among the four groups of speakers, and demonstrated a wider range of gap duration across vowel contexts and clusters than the other three groups of speakers. The average durations of stop gap in the Bi-high and English monolingual children were 100 ms and 110 ms, respectively. The statistical analysis yielded a

TABLE 2. Summary of the statistical results (main effects) of three-way repeated measures ANOVA tests on each durational measurement (absolute duration). For each test, group was set as the between-subject factor, cluster and vowel were set as the within-subject factors

Duration measurement	Group		Cluster		Vowel	
	ANOVA	Post-hoc comparison	ANOVA	Post-hoc comparison	ANOVA	Post-hoc comparison
Cluster	$F(3,26) = 4.898$ $p = .008$ $\eta_p^2 = .361$	Bi-high > EA EC > EA	$F(2,52) = 6.796$ $p = .002$ $\eta_p^2 = .207$	/st/ > /sp/ /st/ > /sk/	$F(2,52) = 32.981$ $p < .0001$ $\eta_p^2 = .559$	/a/ < /i/ /a/ < /u/
Fricative	n.s.	N/A	$F(2,52) = 25.731$ $p < .0001$ $\eta_p^2 = .497$	/st/ > /sp/ /st/ > /sk/	$F(2,52) = 16.009$ $p < .0001$ $\eta_p^2 = .381$	/a/ < /i/ /a/ < /u/
Stop gap	$F(3,26) = 8.150$ $p = .001$ $\eta_p^2 = .485$	Bi-low > EA Bi-high > EA EC > EA	$F(2,52) = 8.103$ $p = .001$ $\eta_p^2 = .238$	/sp/ > /st/ /sp/ > /sk/	$F(2,52) = 4.117$ $p = .022$ $\eta_p^2 = .137$	/a/ < /i/ /a/ < /u/
Stop burst	n.s.	N/A	$F(2,52) = 32.432$ $p < .0001$ $\eta_p^2 = .555$	/st/ > /sp/ /sk/ > /st/ /sk/ > /sp/	$F(2,52) = 3.322$ $p = .044$ $\eta_p^2 = .113$	n.s.
VOT	n.s.	N/A	$F(2,52) = 19.622$ $p < .0001$ $\eta_p^2 = .430$	/st/ > /sp/ /sk/ > /st/ /sk/ > /sp/	$F(2,52) = 13.791$ $p < .0001$ $\eta_p^2 = .347$	/a/ < /i/ /a/ < /u/

TEMPORAL FEATURES OF CONSONANT CLUSTERS

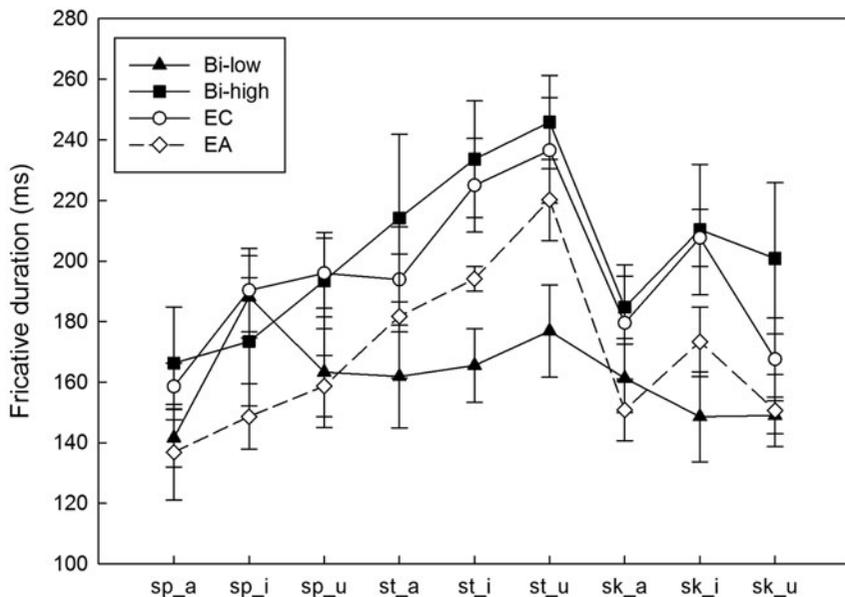


Fig. 2. The mean and standard error of fricative durations for each /s/+stop cluster followed by each of the three vowels /a, i, u/ in each group of speakers.

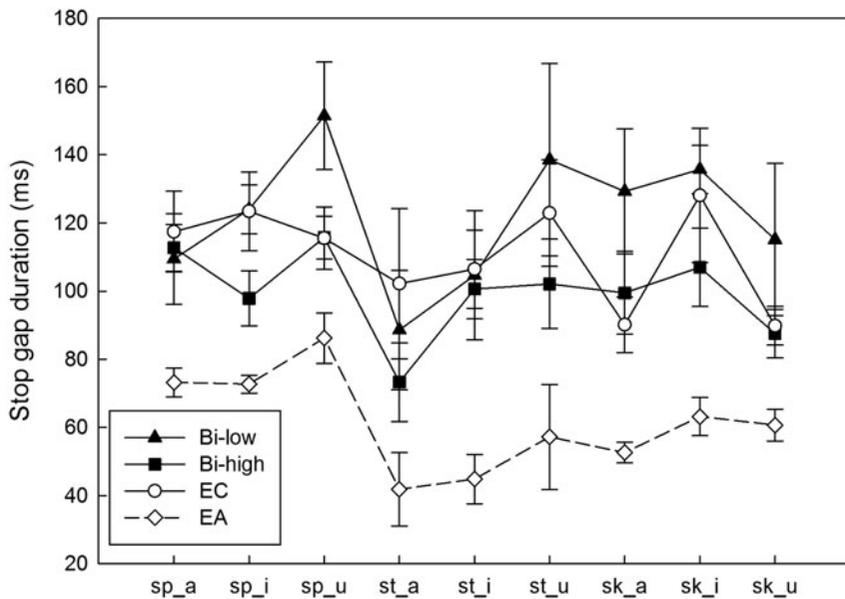


Fig. 3. The mean and standard error of stop gap durations for each /s/+stop cluster followed by each of the three vowels /a, i, u/ in each group of speakers.

significant main effect of group (see [Table 2](#)). The post-hoc analysis revealed that all three groups of children produced significantly longer stop gaps than English monolingual adults, but no differences between the three groups of children were found.

Regarding the VOT and stop burst duration, the four groups of speakers showed similar overall average durations on both measurements. The statistical results revealed a non-significant group effect. However, there were significant effects of vowel and cluster on both measurements (see [Table 2](#)). The variation of VOTs as a function of clusters and vowel contexts was consistent with previous findings (Cho & Ladefoged, 1999; Rochet & Fei, 1991; Weismer, 1979).

Proportional duration of cluster components

[Figure 4](#) presents the means and standard errors of the durational proportions of the fricative to cluster. Averaged across the vowel contexts and stop places, the fricative /s/ weighed 65.3% of the entire cluster in English monolingual adults and only 53.0% in Bi-low children. The Bi-high children and English monolingual children assigned 60.8% and 57.8% of cluster duration on /s/, respectively. This finding indicates that the English monolingual adults allotted more time to the /s/ sound, but the three groups of children, especially the Bi-low children, saved more time for other components when producing the clusters. Compared to the Bi-low children, the Bi-high children showed greater approximation to the English monolingual children and adults. The Bi-high children also demonstrated a pattern of fricative proportions across clusters and vowel contexts more compatible to the English children and adults. The statistical analysis yielded a significant main effect of group (see [Table 3](#)). Post-hoc comparisons revealed that Bi-low children had a smaller proportion of /s/ to cluster than the Bi-high children and the English adults.

[Figure 5](#) demonstrates the means and standard errors of the durational proportions of the stop gap to cluster. Averaged across vowel contexts and stop places, the stop gap weighed 24.2% of the entire cluster in the English monolingual adults and 38.7% in the Bi-low children. For the Bi-high children and English monolingual children, the stop gap weighed 30.9% and 32.6% of the entire cluster, respectively. This finding suggests that the adult speakers used a small portion of time to form the stop closure and to change from the gesture of /s/ to the gesture of the stops. In contrast, the three groups of children, especially the Bi-low children, required more time to do so. Compared to the Bi-low children, the Bi-high children showed a greater approximation to the English monolingual children and adults. The statistical results confirmed the group differences (see [Table 3](#)). The post-hoc analysis showed that the proportions of stop gap

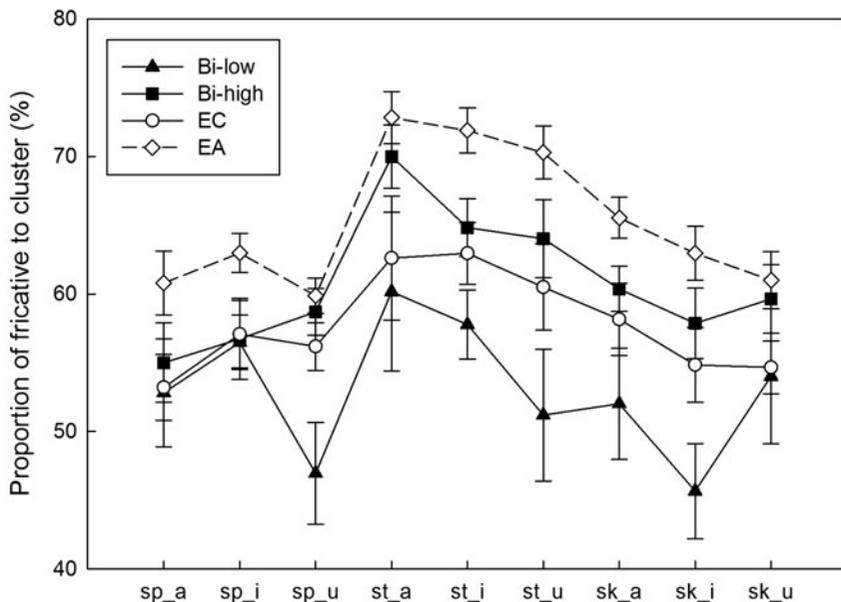


Fig. 4. The mean and standard error of the proportions of fricative duration to cluster duration for each /s/+stop cluster followed by each of the three vowels /a, i, u/ in each group of speakers.

in the Bi-low children were significantly larger than that in the Bi-high children and the English monolingual adults. In addition, the English monolingual children showed greater proportions of stop gap than the English monolingual adults.

As for the durational proportions of the stop burst to cluster, the four groups of speakers showed similar durational proportions of the stop burst, which resulted in no clear group difference. However, a significant main effect of cluster was yielded. This may be associated with the high occurrence of multiple bursts in velar stops. Together with the result of the significantly different absolute duration of stop burst in different clusters, the present study supported the influence of place of articulation on temporal features of stop consonants (Cho & Ladefoged, 1999).

DISCUSSION

The goal of the present study was to investigate how young Mandarin–English bilingual children temporally organize English word-initial /s/+stop clusters in comparison to English monolingual children and adults. Specifically, the absolute duration and durational proportion of the clusters and cluster components were compared between the bilingual

TABLE 3. Summary of the statistical results (main effects) of three-way repeated measures ANOVA tests on each durational measurement (durational proportion). For each test, group was set as the between-subject factor, cluster and vowel were set as the within-subject factors

Duration measurement	Group		Cluster		Vowel	
	ANOVA	Post-hoc comparison	ANOVA	Post-hoc comparison	ANOVA	Post-hoc comparison
Fricative to cluster	$F(3,26) = 6.600$ $p = .002$ $\eta_p^2 = .432$	Bi-low < Bi-high Bi-low < EA	$F(2,52) = 35.478$ $p < .0001$ $\eta_p^2 = .577$	/st/ > /sp/ /st/ > /sk/	$F(2,52) = 3.191$ $p = .049$ $\eta_p^2 = .109$	n.s.
Stop gap to cluster	$F(3,26) = 8.966$ $p < .0001$ $\eta_p^2 = .508$	Bi-low > Bi-high Bi-low > EA EC > EA	$F(2,52) = 32.493$ $p < .0001$ $\eta_p^2 = .556$	/sp/ > /st/ /sp/ > /sk/ /sk/ > /st/	n.s.	N/A
Stop burst to cluster	n.s.	N/A	$F(2,52) = 24.709$ $p < .0001$ $\eta_p^2 = .487$	/st/ > /sp/ /sk/ > /st/ /sk/ > /sp/	$F(4,104) = 2.716$ $p = .034$ $\eta_p^2 = .095$	n.s.

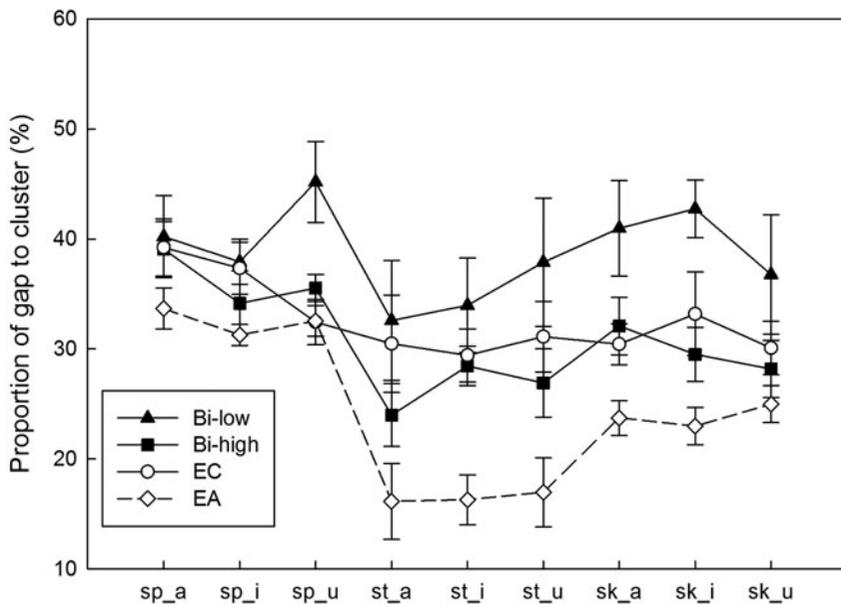


Fig. 5. The mean and standard error of the proportions of stop gap duration to cluster duration for each /s/+stop cluster followed by each of the three vowels /a, i, u/ in each group of speakers.

children and English monolinguals. When we compared bilingual children with English monolingual children, the Bi-high children showed no difference from the English monolingual children in both the absolute durations and durational proportions, which was in agreement with our predictions. Although the Bi-low children tended to produce the English clusters with a longer duration, and assigned a larger proportion of time to the stop gap and a smaller proportion of time to the fricative than the English monolingual children, they did not show a significant difference from the English children. This was inconsistent with our predictions.

When comparing the bilingual children with the English monolingual adults, we found significantly longer cluster duration in the Bi-high children than in the English adults, which was mainly represented in the longer fricative duration and longer stop closure in the Bi-high children. However, the durational proportions of the Bi-high children did not statistically differ from the English monolingual adults. These results suggest that, although the Bi-high children spent more time on the articulatory gestures of individual components in cluster production, they might have developed a strategy of temporal organization similar to the English monolingual adults.

The Bi-low children, inconsistent with our prediction, did not show a significantly longer duration of cluster than the English adults, even though they produced certain components – the stop closure – with longer duration than the English adults. Regarding the durational proportions, the Bi-low children displayed an opposite pattern from the English adults. The English monolingual adults assigned a very small proportion to the stop gap, but a large proportion to the fricative. By contrast, the Bi-low children assigned a relatively small durational proportion to the fricative, but a relatively large durational proportion to the stop closure. These results suggest that it is very likely that the Bi-low children used a distinct strategy of time arrangement in producing the English onset /s/+stop clusters from the English monolingual adults.

We expected that Bi-low children would produce longer durations for the clusters than English monolingual children and adults. However, the results of the present study failed to show evidence to support these predictions. This may be associated with a lack of statistical power due to the relatively small number of participants in each group of participants. In addition, two possible explanations may account for the unexpected results. First, this may be related to the constraint of syllable structure in Mandarin. Mandarin syllables contain single consonants but no consonant clusters. Single consonants are shorter than consonant clusters. When the Mandarin-speaking children started to learn English, and had not mastered the phonetic structure of the new language, their production of L2 sounds was likely to be affected by the temporal constraint in their native language. As a result, the Bi-low children tended to produce the English clusters with short durations, which were similar to those produced by English adults but shorter than those produced by the Bi-high and English monolingual children. Further, because the Bi-low children had not acquired the complex structure of clusters, they produced longer stop closure to transition from the fricative to the following stop. Correspondingly, the fricative /s/ in the Bi-low children was shortened in comparison to the other three groups.

Second, this may be partially associated with the phonetic similarity between Mandarin and English. Both Mandarin and English contain bilabial, alveolar, and velar stops. English stop consonants have a phonological contrast of voiced/voiceless distinction, but they are phonetically represented as short-lag vs. long-lag contrast in most cases, particularly at the syllable-initial position (Keating, 1984). This phonetic representation is similar to the unaspirated/aspirated voiceless contrast in Mandarin. The VOTs of /p, t, k/ in English clusters are similar to those of the unaspirated voiceless stops /p, t, k/ in Mandarin (Chen, Chao & Peng, 2007; Klatt, 1975). It is likely that the Bi-low children transferred the features of the similar speech sounds in their native language to the

new language when they started to learn English. The phonetic similarity of /p, t, k/ between Mandarin and English may also explain why both groups of bilingual children did not show differences in the VOTs from the monolingual English speakers.

Previous research has revealed that children produced many movements more slowly than adults and showed greater temporal variability than adults (Nitttrouer, 1993; Smith & Goffman, 1998). The present study provided further evidence to support this claim. The English monolingual adults produced shorter durations for the entire clusters and cluster elements than the three groups of children. This finding indicates that adults move articulators generally more quickly than children. Moreover, all three groups of children produced longer absolute durations and durational proportions of stop closure than the adults. This indicates that children, regardless of their language experience, tend to spend more time on the articulatory transition from the fricative to the following stop sound in comparison to the adults. In addition, among the four groups of speakers, the Bi-low children produced the stop closure with the longest absolute duration and the largest durational proportion. This observation suggests that Mandarin-speaking children at the initial stage of English learning may experience more difficulty in making articulatory transition from one consonant to the next, which may be partially related to the lack of a complex phonetic structure – consonant blends – in their native language.

In addition to the continuing development of temporal features from children to adults, the present study may provide evidence in favor of the temporal adjustment associated with articulatory coordination during the process of cluster articulation. As presented in Figures 3 and 5, the stop closure in /st/ was shorter than in /sp/ and /sk/. The aerodynamic account suggests that the stop consonant with a small cavity behind the constriction requires less time to build up the intraoral air pressure and thus has a shorter duration for stop closure (Maddieson, 1997). This explanation accounts for the shorter stop closure in /t/ relative to /p/. We assume that the shorter stop closure in /st/ than in the other two may also be related to a short articulatory transition from /s/ to /t/, as these two segments share the same place of articulation. Corresponding to the short stop closure in /st/, the preceding fricative in /st/ was lengthened compared to those in /sp/ and /sk/ (see Figures 2 and 4). These observations indicate the interactions among the components in consonant clusters.

Although informative, the present findings were derived from a relatively small sample size in a small group of participants, which should be generalized with caution. A greater number of participants are needed to verify the current findings. In addition, the present study only examined

word-initial /s/+stop clusters, which presents a narrow focus. Future studies on more complex cluster types in different syllable positions from a larger size of participants, especially the bilingual speakers, are needed to provide further insights into the detailed acoustic-phonetic features and articulatory mechanism of consonant clusters in young bilingual children. Moreover, bilingual children's production of consonant clusters in more strictly controlled speech material and connected speech will also be of interest, because this type of investigation helps us understand the effects of speech rate on the temporal coordination in this population.

REFERENCES

- Baker, R. E., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Van Engen, K. J. & Bradlow, A. R. (2011). Word durations in non-native English. *Journal of Phonetics* **39**(1), 1–17.
- Borden, G. J. & Gay, T. (1979). Temporal aspects of articulatory movements for/s/-stop clusters. *Phonetica* **36**(1), 21–31.
- Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics* **24**(2), 209–44.
- Carlisle, R. S. (1988). The effect of markedness on epenthesis in Spanish/English interlanguage phonology. *Issues and Developments in English and Applied Linguistics (IDEAL)* **3**, 15–23.
- Chen, L. M., Chao, K. Y. & Peng, J. F. (2007). VOT productions of word-initial stops in Mandarin and English: a cross-language study. *Proceedings of the 19th Conference on Computational Linguistics and Speech Processing (ROCLING)*. Online: <http://aclclp.org.tw/rocling/2007/PP04.pdf> (last accessed 3 May 2017).
- Cho, T. & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics* **27**(2), 207–29.
- Eckman, F. R. & Iverson, G. K. (1993). Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research* **9**(3), 234–52.
- Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America* **89**(1), 395–411.
- Fudge, E. C. (1969). Syllables. *Journal of Linguistics* **5**(2), 253–86.
- Gilbert, J. H. & Purves, B. A. (1977). Temporal constraints on consonant clusters in child speech production. *Journal of Child Language* **4**(3), 417–32.
- Greenberg, J. H. (1965). Some generalizations concerning initial and final consonant sequences. *Linguistics* **3**(18), 5–34.
- Guion, S. G., Flege, J. E., Liu, S. H. & Yeni-Komshian, G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics* **21**(2), 205–28.
- Hansen, J. G. (2001). Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics* **22**(3), 338–65.
- Hultzén, L. S. (1965). Consonant clusters in English. *American Speech* **40**(1), 5–19.
- Jayaraman, S. (2010). Acquisition of pronunciation of consonant clusters by Arabic speakers of English as a second language. *Sino-US English Teaching* **7**(1), 46–56.
- Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language* **60**(2), 286–319.
- Kirk, C. & Demuth, K. (2005). Asymmetries in the acquisition of word-initial and word-final consonant clusters. *Journal of Child Language* **32**(4), 709–34.
- Klatt, D. H. (1975). Voice onset time, friction, and aspiration in word-initial consonant clusters. *Journal of Speech, Language, and Hearing Research* **18**(4), 686–706.
- Kohler, K. (1968). Modern English phonology. *Lingua* **19**(1), 145–76.

- Levelt, C. C., Schiller, N. O. & Levelt, W. J. (2000). The acquisition of syllable types. *Language Acquisition* 8(3), 237–64.
- Locke, J. L. (1983). *Phonological acquisition and change*. New York: Academic Press.
- Maddieson, I. (1997). Phonetic universals. In J. Laver & W. J. Hardcastle (eds), *The handbook of phonetic sciences*, 619–39. Oxford: Blackwell.
- McLeod, S., Van Doorn, J. & Reed, V. A. (2001a). Consonant cluster development in two-year-olds: general trends and individual difference. *Journal of Speech, Language, and Hearing Research* 44(5), 1144–71.
- McLeod, S., Van Doorn, J. & Reed, V. A. (2001b). Normal acquisition of consonant clusters. *American Journal of Speech-Language Pathology* 10(2), 99–110.
- Menyuk, P. & Klatt, M. (1975). Voice onset time in consonant cluster production by children and adults. *Journal of Child Language* 2(2), 223–31.
- Nguyen, N. (2008). *Interlanguage phonology and the pronunciation of English final consonant clusters by native speakers of Vietnamese*. Unpublished master's thesis, Ohio University.
- Nittrouer, S. (1993). The emergence of mature gestural patterns is not uniform: evidence from an acoustic study. *Journal of Speech, Language, and Hearing Research* 36(5), 959–72.
- O'Shaughnessy, D. (1974). Consonant durations in clusters. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 22(4), 282–95.
- Riazantseva, A. (2001). Second language proficiency and pausing: a study of Russian speakers of English. *Studies in Second Language Acquisition* 23(4), 497–526.
- Riggenbach, H. (1991). Toward an understanding of fluency: a microanalysis of nonnative speaker conversations. *Discourse Processes* 14(4), 423–41.
- Rochet, B. L. & Fei, Y. (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: production and perception. *Canadian Acoustics* 19(4), 105–6.
- Schwartz, M. F. (1970). Duration of /s/ in /s/-plosive blends. *Journal of the Acoustical Society of America*, 47(4B) 1143–1144.
- 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(4), 779–98.
- Smith, A. & Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language, and Hearing Research* 41(1), 18–30.
- Tingley, B. M. & Allen, G. D. (1975). Development of speech timing control in children. *Child Development* 46(1), 186–94.
- Treiman, R. (1986). The division between onsets and rhymes in English syllables. *Journal of Memory and Language* 25(4), 476–91.
- Trofimovich, P. & Baker, W. (2006). Learning second language suprasegmentals: effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28(1), 1–30.
- Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America* 61(3), 846–58.
- Weismer, G. (1979). Sensitivity of voice-onset-time (VOT) measures to certain segmental features in speech production. *Journal of Phonetics*, 7, 197–204.
- Yavas, M. & Someillan, M. (2005). Patterns of acquisition of/s/-clusters in Spanish–English bilinguals. *Journal of Multilingual Communication Disorders* 3(1), 50–5.