ONSET OBSTRUENT CLUSTER PRODUCTION BY JAPANESE-ENGLISH BILINGUALS: CAN CROSS-LINGUISTIC INTERACTION EXPLAIN EVERYTHING?

SHA LIU, KAYE TAKEDA[1]

Abstract. Cross-linguistic interaction has been used to explain features in the speech production by bilingual speakers with two first languages. The present study focuses on the production of onset obstruent clusters in English by Japanese-English bilingual speakers of 2L1s and compare their production with that by monolingual English speakers and by native Japanese speakers who speak English as a second language. The aim is to examine whether the influence of Japanese can explain differences between the monolingual and bilingual speakers in this paper. Statistical analysis results show that the bilingual speakers have their own patterns in the production of these clusters, which is dissimilar to either monolingual speakers or ESL speakers.

Keywords: bilingual, cross-linguistic interaction, English, Japanese, onset obstruent cluster.

1. INTRODUCTION

The production of onset obstruent clusters in English has received increasing attention in the literature in recent decades; however, the focus has mostly been on second language speakers. An onset obstruent cluster in English is composed of a fricative and a stop, which is against the Minimum Sonority Distance Principle (Steriade 1982, Harris 1983, Selkirk 1984, Broselow and Finer 1991, Eckman and Iverson 1993, Duanmu 2002). It is claimed that onset obstruent clusters in English pose difficulty for those who speak English as a second language, especially for those whose native languages do not permit initial onset clusters (see e.g., Broselow 1983; Carlisle 1991, 1992, 1997, 2006; Hansen 2001; Kwon 2006; O'Neal 2015). The majority of works have examined native Portuguese or Spanish speakers, and generally agree on the point that the production of onset obstruent

[1] Fukuoka University (Japan), liusha@fukuoka-u.ac.jp, kayeliou@gmail.com.

For help in getting this article to its final form, our special gratitude goes to Prof. Jacques Durand for his advice on acoustic analysis, to Prof. Eiji Yamada and Prof. Hajime Takeyasu for advice and discussion, to Prof. Greg Bevan and Prof. Robert Long for editing our paper, and to the anonymous reviewer and the editors of the present journal for detailed and helpful feedback. All remaining errors are our responsibility. This work was funded by JSPS Grant-in-Aid for Early-Career Scientists (KAKENHI-PROJECT-20K13072).

clusters by these speakers is somewhere between their respective native languages and English. This has been explained as the co-activation of their respective native languages (see e.g., Broselow and Finer 1991, Carlisle 1991, Eckman and Iverson 1993, Chan 2010).

Despite increased interest in bilingual speakers, scholars still cannot agree with each other on whether cross-linguistic interaction between the two language systems and structures in bilinguals exists (see e.g., Caramazza et al. 1973, Major 1987, Hazan and Boulakia 1993, Kroll et al. 2012, Simonet 2014). Since previous studies have reported a generally unified result concerning the production of onset obstruent clusters in English by ESL speakers, it appears plausible that the production of these clusters by bilingual speakers may present hints concerning differences, if any, between monolingual and bilingual speakers. The present study attempts to investigate the production of onset obstruent clusters in English by Japanese-English bilinguals with two first languages since obstruent clusters do not exist in the onset position in Japanese and very few previous studies have examined Japanese-English bilingual speakers from this perspective. This study also takes monolingual English speakers and native Japanese speakers who speak English as a second language (hereafter Japanese ESL speakers) into consideration since these speakers provide a benchmark for comparison with bilingual speakers. Given the context of situation as briefly described here, the first question that this paper addresses is whether differences exist in the production of onset obstruent clusters by monolingual English speakers and Japanese-English bilingual speakers. If the answer to this question is affirmative, the paper further explores whether cross-linguistic interaction can sufficiently account for these differences.

The present study is structured as follows. Section 2 reviews the results for the production of onset obstruent clusters in English in previous studies and then provides associated hypotheses. Section 3 details the experiment for this study while Section 4 statistically analyzes the results of the experiment; Section 5 concludes the whole paper.

2. PREVIOUS STUDIES

In the following, we firstly give a description of onset obstruent clusters in English. Secondly, we present the difficulty these clusters in English pose for ESL speakers as discussed in previous studies. Thirdly, we discuss the concept of cross-linguistic interaction in previous studies and disagreements surrounding it. Finally, we propose our own hypothesis concerning Japanese-English bilingual speakers.

2.1. Onset obstruent clusters in English

The most common syllable structure in English is CVC (McCully 2009). In addition, up to three consonants are allowed in the onset position and four consonants in the coda position in English (Dauer 1983; Shockey 2003; Schreier 2005; McCully 2009; McLeod 2010). Among all the permissible syllable structures in English, onset obstruent clusters in English have a special status as they are exceptions to the Sonority Sequencing Principle (henceforth SSP; Broselow 1983, Selkirk 1984, Yavaş and Someillan 2006). Sonority is "a unique type of relative, n-ary (non-binary) feature … that potentially categorizes all

speech sounds into a hierarchical scale" (Parker 2011: 1160). One major function of the SSP is to organize segments in syllables: SSP suggests that the segment ranking highest on the sonority scale should be the peak of a syllable; all other segments are organized around the nucleus in a way that the more sonorous segments are closer to the peak and the less sonorous ones are further away from it (Sievers 1881, Jespersen 1904, Saussure 1916, Grammont 1933, Steriade 1982, Harris 1983, Clements 1990). Although disputes surrounding the SSP are not rare, its role as a most common restriction on segment sequencing is undeniable. Different scholars have proposed different versions of the sonority scale (Jespersen 1904, Saussure 1916, Grammont 1933, Hogg and McCully 1987, Clements 1990). The most widely accepted version is perhaps the five-point sonority scale as follows (Clements 1990, Kenstowicz 1994, Smolensky 1995).

(1) The five-point sonority scale (Clements 1990)
vowels > glides > liquids > nasals > obstruents

In (1), the mark > means 'more sonorous than'. Thus the five-point sonority scale in (1) decreases in sonority from the left to the right by an interval of one. To exemplify, the sonority scale in (1) indicates that vowels are more sonorous than glides; nasals are more sonorous than obstruents. In addition to the SSP, the Minimum Sonority Distance (hereafter MSD) is also proposed, which is "a requirement on the minimal sonority distance … between the two sounds in an onset cluster" (Duanmu 2002: 99; see also Steriade 1982, Harris 1983, Selkirk 1984, Broselow and Finer 1991, Eckman and Iverson 1993). In English, the MSD is assumed to be two (Duanmu 2002, Ostapenko 2005). For example, an obstruent-liquid cluster in the onset position in English is thought to be better than an obstruent-nasal cluster since the sonority distance between the two components in an obstruent-liquid cluster is two and the sonority distance between the two components in an obstruent-nasal cluster is only one.

Well-known counterexamples to MSD include obstruent clusters at the onset position in English, such as /sk/, /sp/, and /st/. The sonority distance between the two components in these clusters is zero since the two elements in each cluster are both obstruents. In other words, the sonority does not rise from /s/ to the neighboring stop in each cluster. One note needs to be made before we leave this section. We follow the sonority scale in Clements (1990) in this paper. Although differences exist between different versions of the sonority scale, we have not found any proposal in which fricatives are less sonorous than stops. To exemplify, Hogg and McCully (1987) present a ten-point sonority scale, in which the fricatives are considered more sonorous than the stops. Therefore, their version of sonority scale still suggests that onset obstruent clusters in English are against MSD.

2.2. Difficulties for ESL speakers

Previous studies have mainly focused on the difficulty of onset obstruent clusters in English for ESL speakers whose native languages do not allow onset obstruent clusters. A large number of scholars take native speakers of Portuguese or Spanish as their focus (see e.g., Carlisle 1991, 1992, 1997, 2006; Rauber and Baptista 2004; Escartín Ortiz 2005). Other scholars focus on native Arabic speakers (see e.g., Broselow 1983, Jayaraman 2010),

native Japanese speakers (see e.g., Broselow and Finer 1991, O'Neal 2015), native Korean speakers (see e.g., Kim 2000, Kwon 2006), native Mandarin speakers (see e.g., Hansen 2001), and so on.

Co-activation of a first language has been invoked to explain the difficulty involved in the production of English onset obstruent clusters by ESL speakers (Lado 1957, Eckman and Iverson 1993). Co-activation means that the comparison and contrast of a first language and a second language is related to the acquisition of the second language (Lado 1957, Moulton 1962, Stockwell and Bowen 1965, Odlin 1989). To exemplify, Carlisle (1991) reports that onset obstruent clusters in English are difficult for native Portuguese and Spanish speakers. He (1991) reports that speakers make more mistakes with onset clusters against MSD than onset clusters with a large sonority distance (see also Eckman and Iverson 1993, Cardoso and Liakin 2009). The most common strategy for ESL speakers is internal vowel epenthesis (Rauber and Baptista 2004, Kelly 2014; see a similar strategy for Farsi speakers reported in Boudaoud and Cardoso 2009). This results in the syllabification of an obstruent sequence into two syllables. A related example is the Latin word *stabula* 'steady' which became the Spanish word *es.table*. The insertion of the vowel /e/ at the beginning of this Latin word syllabifies the cluster /st/ into two syllables. Deletion of the second consonant in a cluster is rare. Similarly, Chan (2010) takes native Cantonese speakers as her focus and examines their production of English obstruent-nasal onsets and obstruent-liquid onsets. She (2010) also claims that the difficulty is mainly caused by the phonotactic constraints of these learners' first language. Albeit the lack of complete agreement with each other, most scholars agree that the production of onset obstruent clusters by ESL speakers is somewhere between their first language and English. To exemplify, Broselow and Finer (1991) examine the production of onset clusters, e.g., /br/, /fr/, /pr/, in English by native speakers of Hindi, Japanese, and Korean, and report that their results suggest that the realization of these clusters by these ESL speakers is somewhere between their respective native languages and the target language English, and co-activation of their first languages to the second language English has been witnessed (see also Eckman and Iverson 1993).

2.3. Bilinguals: Intermediate between monolinguals and second language speakers?

Recent years have also seen an emerging interest in speech production by bilingual speakers. Caramazza et al. (1973) claim that bilingual speakers establish and produce separate phonetic categories for their two languages (see also e.g., Hazan and Boulakia 1993, MacLeod and Stoel-Gammon 2010). More recently, Paradis and Genesee (1996) focus on French-English bilingual children and find that their acquisition of finiteness, negation and subject pronouns in French and English is respectively the same as monolingual peers (see also De Houwer 1995, Meisel 2004). Magloire and Green (1999) ask Spanish-English bilinguals to produce sentences containing voiced and voiceless bilabial stops at different speaking rates in Spanish and English respectively, and find that their VOT values of voiced and voiceless stops are largely equivalent to respective monolingual peers. Thus, their study has provided evidence for separate categories, at least separate categories for the bilabial stops in the two languages in the Spanish-English bilinguals. Antoniou et al. (2010) research into Greek-English bilinguals and their VOT values in both Greek and English contexts, and come to a similar conclusion as Magloire

and Green (1999). Engemann (2021) focuses on simultaneous English-French bilingual children but comes to a less controversial conclusion, claiming that the motion-event encoding tendencies of the bilingual children are considerably closer to those of their monolingual peers (see also Nathan et al. 1987, Hazan and Boulakia 1993, Macleod and Stoel-Gammon 2005).

On the other hand, a growing body of research supports cross-linguistic interaction between the two language systems and structures in bilinguals: both languages are constantly and jointly active to some degree even when only one language is being used (Van Hell and Dijkstra 2002; Von Studnitz and Green 2002; Dijkstra 2007; Wu and Thierry 2010a, 2010b; Kroll et al. 2012; Simonet 2014; see also the review in Hambly et al. 2013). In other words, the two languages are integrated in bilingual speakers. For example, Nicoladis (2006) finds that French-English bilingual children reversed more adjective-noun strings in English than monolinguals (see also Hsin et al. 2013, and the review in Serratrice 2013). Filipović (2011) focuses on balanced English-Spanish bilingual speakers and claims that these speakers tend to use a single lexicalization pattern that is acceptable in both English and Spanish to describe and remember complex motion events (see also Flege and Port 1981, Major 1987, Flege and Eefting 1987).

In a word, there are still disagreements concerning whether cross-linguistic interaction exists in bilinguals. Since previous studies have generally agreed that the production of onset obstruent clusters by ESL speakers is somewhere between English and their native languages, it appears that a comparison of the production of onset obstruent clusters in English between monolingual speakers, bilingual speakers, and ESL speakers could provide further evidence concerning cross-linguistic interaction or lack thereof in bilingual speakers.

2.4. The proposal of the present study

The most common syllable structure in Japanese is CV (Vance 1987; Riney and Anderson-Hsieh 1993; Kubozono 1999). Japanese generally has no consonant clusters (Riney and Anderson-Hsieh 1993), although disagreements exist over whether palatalized consonants are single consonants or consonant clusters. Some researchers treat them as /Cj/ which consists of a consonant followed by a glide (Broselow and Finer 1991, Nogita 2016, Hirayama and Vance 2018), while others claim that they are a single phoneme /C^{j}/ (Ito and Mester 1995, Ohata 2004). Returning to the present study, it is unanimously agreed that onset obstruent clusters, such as /sk/, /sp/, and /st/, do not exist in Japanese. English consonant sequence acquisition and production has been found to be difficult for Japanese speakers due to the lack of initial and final consonant clusters in Japanese (Ohata 2004, Smith 2006, Takeshita 2010, Scherling 2012). Vowel insertion is a common strategy used by Japanese speakers: vowels tend to be added to break up consonant clusters in a word, which results in a CV pattern in the word (Kenworthy 1987, Thompson 1987, Weinberger 1987, Smith 2006, Scherling 2012). Based on these features in Japanese, this paper focuses on Japanese-English bilinguals and their production of onset obstruent clusters in English. To have a benchmark for a clear comparison, this paper includes three groups of speakers: a monolingual group of English, a Japanese-English bilingual group, and a Japanese ESL group.

The monolingual English speakers in this paper are defined as native speakers of English who only command English: they cannot produce effective speech in another

language; nor can they comprehend another language (see e.g., Snow and Hakuta 1992, Mack 1997, Ellis 2007). The bilingual speakers in this paper are defined as those who have acquired both Japanese and English in their infancy and can produce fluent and effective speech in both languages (see e.g., Haugen 1953, Weinreich 1953). As these participants acquired both languages before they were three years old, they are 2L1 speakers (Leopold 1970, Genesee and Nicoladis 2006). Our definition of a bilingual speaker is not as strict as that of Bloomfield (1933) as a perfect user of two languages. However, it is much stricter than that of MacNamara (1967) who includes anyone who has a minimal competence in listening, reading, speaking, or writing a language other than his/her native language. The Japanese ESL speakers in this paper are native speakers of Japanese who did not learn English in their early childhood, received English education at school, and have not lived in any English-speaking country for longer than one month (see e.g., Jenkins 2000, Mitchell and Myles 2004, Kormos 2006). These speakers have an intermediate proficiency in English.

The present paper intends to measure the duration of each component in a cluster, the duration of each onset obstruent cluster, and the durational ratio of the two components in a cluster. The duration of each component in a cluster reflects the amount of time a participant used for gestural movement of individual elements. The duration of each cluster can provide a general depiction of differences between every two groups. The durational ratio is defined as the relative weight of the two elements in a cluster. This measure shows the relationship between one component and the other in each cluster, and thus gives a hint about temporal organization in the production of onset obstruent clusters by each group. Along the discussion, intra-group variations are also examined.

Our general hypothesis is that the bilingual group in this study has intermediate results between the monolingual group and the ESL group in terms of all measures based on the assumption of cross-linguistic interaction between the two language systems and structures in bilinguals following Dijkstra et al. (1998) and Van Hell and Dijkstra (2002) (see also Von Studnitz and Green 2002; Wu and Thierry 2010a, 2010b; Kroll et al. 2012; Simonet 2014). Based on our subjective perception of speech production by bilingual speakers, we further hypothesize that the results from the bilingual group are generally closer to the results from the monolingual group than to those from the ESL group. We accordingly expect statistically significant differences to emerge between the monolingual and bilingual groups in terms of one or two measures, but not in terms of all or most of the measures. In addition, we expect intra-group differences in consideration of the generally consistent results concerning the existence of intra-group differences in previous studies (see e.g., Zhang et al. 2006, Gerosa et al. 2007, Albrecht 2017, Kim et al. 2017, Entringer 2021).

3. EXPERIMENT

Three monolingual native speakers of English, one male and two female, were recorded. The three monolingual native speakers of English (hereafter the monolingual group) were born and brought up in California. They were also residents of California at the time of recording. They cannot communicate effectively in any language other than English. Three Japanese-English bilingual speakers (hereafter the JE bilingual group), one male and two female, were recorded. They were all born in the western part of Japan, moved to

California in their infancy, and spent their formative years in California. Japanese is the language they speak at home. They were also residents of Japan at the time of recording. They acquired Japanese and English naturalistically and are fluent in both Japanese and English. In other words, they are early bilinguals with two first languages (2L1s). We tried to enroll JE bilingual speakers among Californian residents, but it was difficult to recruit three JE bilingual speakers in California. We limited speakers in the monolingual group and the JE bilingual group to Californian English speakers to reduce potential influences of different English accents as much as possible. Three Japanese ESL speakers (hereafter the Japanese ESL group), one male and two female, were also recorded. They were all born and brought up in the western part of Japan. They did not learn English in their early childhood, received English education at school, and have not lived in any English-speaking country for more than one month. Their accents are not notably different from standard Japanese. All speakers in the Japanese ESL group are from the same area of Japan. The aim is also to reduce possible influences of different Japanese accents.

All monolingual speakers and JE bilingual speakers were around 30 to 35 years old at the time of recording and college graduates. All Japanese ESL speakers were college students and were just over 20 years old at the time of recording. Although the Japanese ESL speakers are not of exactly the same age as the monolingual speakers and the JE bilingual speakers, the age difference is not considerably large.

3.1. Recording

All speakers were given the text from the PAC project, *Christmas Interview of a Television Evangelist*, three weeks before their recordings to become familiar with it.[2] The text is embedded with a number of interesting phonological phenomena, including the focus of this paper. We chose text reading instead of word list reading because we wanted to elicit spontaneous pronunciation since participants may make conscious correction or hypercorrection while reading word lists. The participants were instructed to practice reading the passage in their normal voice and at a rate that they felt natural and comfortable until they could read the passage fluently. They were also instructed that they should repeat the whole sentence if they made a mistake. The ideal place to record appears to be a sound-proof room at our university. However, the physical distance prevented this option. In addition, recent studies have shown that smartphone recording is acceptable for acoustic analysis (see e.g., Maryn et al. 2017, Oliveira et al. 2017, Wu 2017, van der Woerd et al. 2020, Uloza et al. 2021). Thus, recordings were made on the second author's iPhone 6 in quiet rooms. The original format of the recordings was m4a. They were later converted to the wav format for acoustic analysis on Praat.

3.2. Segmentation and analysis

The first author segmented and labeled the recorded speeches. This procedure was carried out on speech waveforms and wideband spectrograms generated on Praat. The guidelines laid out in Peterson and Lehiste (1960), Grabe and Low (2002), and White and

[2] *PAC* stands for *La Phonologie de l'Anglais Contemporain: Usages, Variétés et Structure* in French or *The Phonology of Contemporary English: Usage, Varieties and Structure* in English.

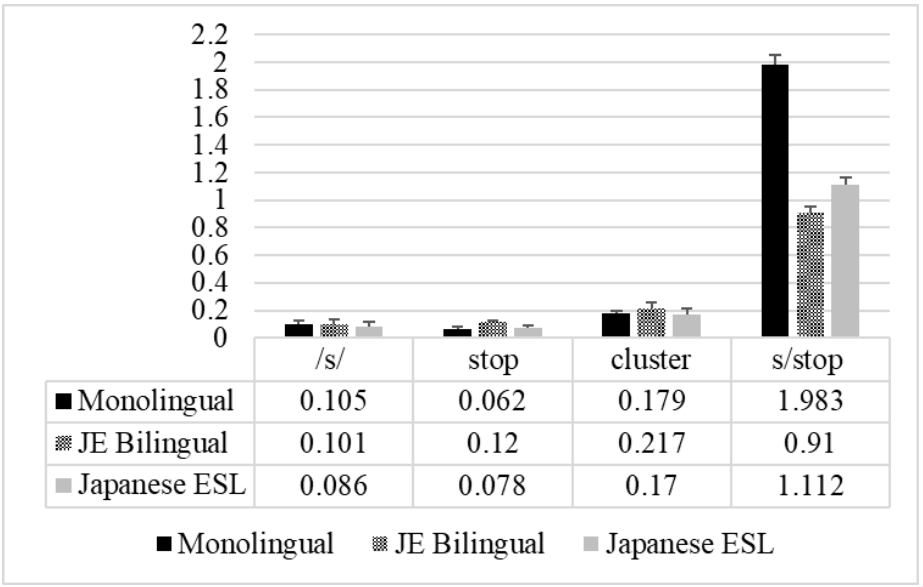
Mattys (2007) were generally followed. Pauses between intonation phrases were excluded from the analysis. The segmental boundaries were generally identified by taking spectral transitions into consideration. The boundary between an onset obstruent cluster and the following vowel was the beginning of the pitch period at the onset of the vocalic formant structure. The boundary between /s/ and a following stop was the beginning of a silent stop gap shown on the spectrum. Forty-eight samples were collected from each group, a total of 144 samples. The clusters focused on are /sk/, /sp/, and /st/.

4. EXPERIMENT RESULTS

Statistical analysis was carried out on the software GraphPad Prism Version 8.0.0 for Windows (GraphPad Software, Inc., San Diego, CA; hereafter GraphPad software). Details are given in the following two subsections.

4.1. The four measures

As noted in Section 2.4, we focus on the duration of each component in a cluster, the duration of each cluster, and the durational ratio of the two components in a cluster. The mean value and standard deviation for each group for each of these measures are given in Figure 1.



	/s/	stop	cluster	s/stop
Monolingual	0.105	0.062	0.179	1.983
JE Bilingual	0.101	0.12	0.217	0.91
Japanese ESL	0.086	0.078	0.17	1.112

Figure 1.

The mean value and standard deviation for each group in terms of each measure.
The vertical bars stand for the mean value of each measure for each group. The error bars stand for the standard deviation in terms of each measure for each group.

The marks */s/*, *stop*, and *cluster* in Figure 1 respectively refer to the duration of the /s/ component in a cluster, the duration of the stop component in a cluster, and the duration of a cluster. The mark *s/stop* stands for the durational ratio of these two components in a cluster. Figure 1 graphically shows that only one result is completely in line with our hypothesis: the JE bilingual group has the intermediate mean value in terms of the /s/ duration among the three groups. In addition, its mean /s/ value is much closer to that of the monolingual group than to that of the Japanese ESL group. The JE bilingual group has the highest mean values in terms of the stop duration and cluster duration, whereas the monolingual and Japanese ESL groups have more similar results for these two measures. The JE bilingual group has the lowest ratio of 0.91 in terms of the s/stop ratio, which is close to the ratio of 1.112 for the Japanese ESL group. In contrast, the monolingual group has the highest ratio of 1.983. These results indicate that there is a large difference in the temporal organization of the two elements in an onset obstruent cluster by the JE bilingual and monolingual groups: while the element /s/ has a roughly similar duration to the stop element in the JE group, the element /s/ takes up most of the cluster duration in the monolingual group. As noted earlier, the JE bilingual group has a similar result in terms of the /s/ duration to the monolingual group. Thus the large difference in the s/stop ratio measure is mainly due to the difference in terms of the stop duration between the two groups. The results in Figure 1 show that the production of the onset obstruent clusters in English by the JE bilingual speakers is not the same as the production by the monolingual speakers. However, the results here do not indicate that both languages are jointly active to some degree in the bilingual speakers since our results do not show many similarities between the JE bilingual and ESL groups either. Figure 1 shows that the two groups only share similar results in terms of the s/stop ratio measure and the JE bilingual group has the most extreme results in terms of the stop duration, the cluster duration, and the s/stop ratio measures. In other words, our results appear to indicate that the bilingual speakers have their own characteristics that cannot be explained by cross-linguistic interaction: the JE bilinguals appear to have a temporal organization for the onset obstruent clusters in English different from both the monolingual and ESL speakers.

Brown-Forsythe ANOVA Tests carried on the GraphPad software showed statistically significant differences among the monolingual group, the bilingual group, and the Japanese ESL group in terms of the /s/ duration ($F^*(2, 116.5) = 5.98$; $p = 0.00$), significant differences among the three groups in terms of the stop duration ($F^*(2, 92.98) = 119.5$; $p < 0.00$), significant differences among the three groups in terms of the cluster duration ($F^*(2, 116.1) = 22.88$; $p < 0.00$), and significant differences among the three groups in terms of the s/stop ratio ($F^*(2, 67.08) = 35.13$; $p < 0.00$), where p smaller than 0.05 is taken as statistically significant. Post hoc Dunnett's T3 multiple comparisons test was also carried out on the same software, whose results are reported as follows.

Table 1

Statistical analysis results

	Monolingual vs. Bilingual	Bilingual vs. ESL	Monolingual vs. ESL
/s/	$p = 0.83$	$p = 0.08$	$p = 0.00$
stop	$p = 0.00$	$p = 0.00$	$p = 0.00$
cluster	$p = 0.00$	$p = 0.00$	$p = 0.38$
s/stop	$p = 0.00$	$p = 0.13$	$p = 0.00$

Table 1 shows that, in terms of /s/ duration, significant differences have emerged between the monolingual and ESL groups (p = 0.00), but not between the monolingual and bilingual groups (p = 0.83), or between the bilingual and ESL groups (p = 0.08). Statistically significant differences occur between every two groups in terms of the stop duration as shown by the p values in the third row (p = 0.00, 0.00, 0.00). Results in terms of the cluster duration appear to show that bilingual speakers have their unique features and patterns: differences between the monolingual and ESL groups are not statistically significant (p = 0.38), while the other two combinations show statistically significant differences (p = 0.00, 0.00). In terms of the s/stop ratio, the bilingual group shows a more similar result with the ESL group (p = 0.13), while the monolingual and bilingual groups, and the monolingual and ESL groups exhibit statistically significant differences (p = 0.00, 0.00).

Figure 1 and Table 1 together indicate that the differences between the monolingual and JE bilingual groups are not as small as we have expected. The second column of Table 1 shows that the two groups are similar only in one respect, the /s/ duration (p = 0.83). In terms of the stop duration, cluster duration, and s/stop ratio, the two groups show statistically significant differences as indicated by the p values (p = 0.00, 0.00, 0.00). A comparison between the bilingual and ESL groups further confirms our conclusion: the two groups have statistically significant differences in terms of the stop duration and cluster duration (p = 0.00, 0.00). Both Figure 1 and the statistical analysis results in Table 1 have shown that the monolingual and bilingual groups have similar mean /s/ duration but different mean stop and cluster durations, while the monolingual and ESL groups have similar mean cluster duration but different mean /s/ and stop durations. In other words, this shows that the three groups have different temporal patterns.

4.2. Intra-group differences

The ordinary one-way ANOVA performed on the GraphPad software demonstrates intra-group differences in our data. The three speakers in the monolingual group show statistically significant differences in terms of the stop duration, cluster duration, and s/stop ratio (p = 0.01, 0.00, 0.04), but not in terms of the /s/ duration (p = 0.17). The post hoc Holm-Sidak's multiple comparisons test shows statistically significant differences between one comparison (p = 0.01), but not between the other two comparisons (p = 0.06, 0.32) in terms of stop duration. A similar result was obtained in terms of s/stop ratio (p = 0.04, 0.11, 0.55). In terms of cluster duration, two comparisons were statistically significant (p = 0.05, 0.00), while the last one was not (p = 0.08).

The three speakers in the JE bilingual group display statistically significant results in terms of stop duration and s/stop ratio, but not in terms of /s/ duration and cluster duration (p = 0.02, 0.01, 0.31, 0.15). The post hoc Holm-Sidak's multiple comparisons test shows statistically significant differences between one comparison (p = 0.02), but not between the other two comparisons (p = 0.74, 0.83) in terms of stop duration. In terms of s/stop ratio, two comparisons emerged as statistically significant (p = 0.01, 0.03), while the last one did not (p = 0.82).

The speakers in the ESL group exhibit significant intra-group differences in terms of stop duration and cluster duration (p = 0.04, 0.03). The post hoc Holm-Sidak's multiple comparisons test shows statistically significant differences between one comparison (p = 0.05), but not between the other two comparisons (p = 0.22, 0.44) in terms of stop duration. In

terms of cluster duration, one comparison emerged as statistically significant ($p = 0.04$), the other two did not ($p = 0.35, 0.35$).

In summary, participants in each group share similarities in certain aspects: no group has shown intra-group differences in terms of all the four measures; nor has one group shown intra-group differences in all the three comparisons in terms of one measure. This appears to indicate that although intra-group variation can be a challenge, the pooled data from each group show a general trend toward the same direction.

5. CONCLUSION

This study has focused on Japanese-English bilingual speakers and their production of onset obstruent clusters in English. Our results have demonstrated that bilingual speakers are not simply equivalent to the two monolinguals. They have their own features and patterns, which cannot be satisfactorily explained from the perspective of cross-linguistic interaction. The bilingual group has the most extreme values in terms of three measures out of the four measures we examined. This does not appear to be explicable from the perspective of cross-linguistic interaction: if the influence from Japanese is the main reason here, the JE bilingual group should not have such extreme results for these measures and should have more similarities with the Japanese ESL group. While it may be too premature to give any conclusive answer after examining one aspect of speech production, our data do suggest hints towards the direction that bilingual speakers have their own unique patterns, at least in certain linguistic respects.

The conclusion for this paper was drawn from a relatively small number of participants, so caution is necessary in the overall interpretation of its results. Whether a similar conclusion can be drawn with a larger number of subjects is also a question that needs further research.

REFERENCES

Albrecht, K. M., 2017, *The identification of typical and atypical phonological acquisition in Turkish-German bilingual children*, Doctoral dissertation, The University of Sheffield, Sheffield.

Antoniou, M., C.T. Best, M.D. Tyler, C. Kroos, 2010, "Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2", *Journal of Phonetics*, 38, 4, 640–653. https://doi.org/10.1016/j.wocn.2010.09.005.

Bloomfield, L., 1933, *Language*, New York, Holt, Rinehart & Winston.

Boudaoud, M., W. Cardoso, 2009, "Vocalic [e] epenthesis and variation in Farsi-English interlanguage speech", *Concordia Working Papers in Applied Linguistics*, 2, 1–34.

Broselow, E., 1983, "Non-obvious transfer: On predicting epenthesis errors", in S.M. Gass, L. Selinker (eds), *Language transfer in language learning*, Rowley, MA, Newbury House, 269–280.

Broselow, E., D. Finer, 1991, "Parameter setting in second language phonology and syntax", *Interlanguage studies bulletin (Utrecht)*, 7, 1, 35–59.

Caramazza, A., G.H. Yeni-Komshian, E.B. Zurif, E. Carbone, 1973, "The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals", *Journal of the Acoustical Society of America*, 54, 2, 421–428.

Cardoso, W., D. Liakin, 2009, "When input frequency patterns fail to drive learning: Evidence from Brazilian Portuguese English", in M.A. Watkins, A.S. Rauber, B.O. Baptista (eds), *Recent research in second language phonetics/phonology: Perception and production*, Newcastle Upon Tyne, Cambridge Scholars, 174–202.

Carlisle, R.S., 1991, "The influence of environment on vowel epenthesis in Spanish/ English interphonology", *Applied Linguistics*, 12, 1, 76–95.

Carlisle, R.S., 1992, "Environment and markedness as interacting constraints on vowel epenthesis", in J. Leather, A. James (eds), *New Sounds 92: Proceedings of the 1992 Amsterdam Symposium on the Acquisition of Second-language Speech*, University of Amsterdam, 64–75.

Carlisle, R.S., 1997, "The modification of onsets in a markedness relationship: Testing the interlanguage structural conformity hypothesis", *Language Learning*, 47, 2, 327–361.

Carlisle, R.S., 2006, "The sonority cycle and the acquisition of complex onsets", in B.O. Baptista, M.A. Watkins (eds), *English with a Latin beat*: *Studies in Portuguese*/*Spanish-English* interphonology, Amsterdam, John Benjamins, 105–138.

Chan, A.Y.W., 2010, "An investigation into Cantonese ESL learners' acquisition of English initial consonant clusters", *Linguistics*, 48, 1, 99–141.

Clements, G.N., 1990, "The role of the sonority cycle in core syllabification", in J. Kingston, M.E. Beckman (eds), *Papers in laboratory phonology I: Between the grammar and physic of speech*, New York, Cambridge University Press, 282–333.

Dauer, R.M., 1983, "Stress-timing and syllable-timing reanalyzed", *Journal of Phonetics*, 11, 51–62.

De Houwer, A., 1995, "Bilingual language acquisition", in P. Fletcher, B. MacWhinney (eds), *The handbook of child language*, Cambridge, Blackwell, 219–250.

Dijkstra, T., H. van Jaarsveld, S.T. Brinke, 1998, "Interlingual homograph recognition: Effects of task demands and language intermixing", *Bilingualism: Language and Cognition*, 1, 51–66.

Dijkstra, T., 2007, "The multilingual lexicon", in M. Garet Gaskell (ed), *Handbook of Psycholinguistics*, Oxford, Oxford University Press, 251–265.

Duanmu, S., 2002, "Two theories of onset clusters", *Chinese Phonology*, 11, 97–120.

Eckman, F.R., G.K. Iverson, 1993, "Sonority and markedness among onset clusters in the interlanguage of ESL learners", *Second Language Research*, 9, 3, 234–252.

Ellis, E., 2007, "Monolingualism: The Unmarked Case", *Estudios de Sociolingüística*, 7, 2, 173–196.

Engemann, H., 2021, "How (not) to cross a boundary: Crosslinguistic influence in simultaneous bilingual children's event construal", *Bilingualism: Language and Cognition*, 25, 1, 1–13. https://doi.org/10.1017/ S1366728921000298.

Entringer, N., 2021, "Inter- and intra-individual variation in Luxembourgish: A quantitative analysis of crowd-sourced speech data", in L. Bülow, S.E. Pfenninger, M. Schiegg, A. Werth (eds), *Intra-individual variation in language*, Berlin, De Gruyter Mouton, 243–282.

Escartín, O., C. Ivette, 2005, *The development of sC onset clusters in Spanish English*, Master's thesis, Concordia University.

Filipović, L., 2011, "Speaking and remembering in one or two languages: Bilingual vs. monolingual lexicalization and memory for motion events", *International Journal of Bilingualism*, 15, 4, 66–485.

Flege, J.E., R. Port, 1981, "Cross-language phonetic interference: Arabic to English", *Language and Speech*, 24, 125–146.

Flege, J., E., Wieke, 1987, "The production and perception of English stops by Spanish speakers of English", *Journal of Phonetics*, 15, 67–83.

Genesee, F., E. Nicoladis, 2006, "Bilingual first language acquisition", in E. Hoff, M. Shatz (eds), *Blackwell handbook of language development*, West Sussex, Wiley Blackwell, 324–342.

Gerosa, M., D. Giuliani, F. Brugnara, Fabio, 2007, "Acoustic Variability and Automatic Recognition of Children's Speech", *Speech Communication*, 49, 10–11, 847–860.

Grabe, E., E.L. Low, 2002, "Durational Variability in Speech and the Rhythm Class Hypothesis", in C. Gussenhoven, N. Warner (eds), *Laboratory Phonology* 7, Berlin, Mouton de Gruyter, 515–546.

Grammont, M., 1933, *Traite de phonétique*, Paris, Delagrave.
GraphPad Prism, Version 8.0.0 for Windows, San Diego, California, GraphPad Software, 2019, Computer software.
Hambly, H., Y. Wren, S. McLeod, S. Roulstone, 2013, "The influence of bilingualism on speech production: A systematic review", *International Journal of Language & Communication Disorders*, 48, 1, 1–24.
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.
Harris, J., 1983, *Syllable structure and stress in Spanish: A nonlinear analysis*, Cambridge, Mass., MIT Press.
Haugen, E., 1953, *The Norwegian language in America: A study of bilingual behavior*, Philadelphia, University of Pennsylvania.
Hazan, V.L., G. Boulakia, 1993, "Perception and production of a voicing contrast by French-English bilinguals", *Language and Speech*, 36, 17–38.
Hirayama, M., T.J. Vance, 2018, "Onset Cy and high vowel devoicing in Japanese", *Journal of Japanese Linguistics*, 34, 1, 103–126.
Hogg, R.M., C. McCully, 1987, *Metrical phonology: A course book*, New York, Cambridge University Press.
Hsin, L., G. Legendre, A. Omaki, 2013. "Priming cross-linguistic interference in Spanish-English bilingual children", in S. Baiz, N. Goldman, R. Hawkes (eds), *Proceedings of the 37th Annual Boston University Conference on Language Development*, Somerville, MA, Cascadilla Press, 165–177.
Ito, J., A. Mester, 1995, "Japanese Phonology", in John A. Goldsmith (ed), *The Handbook of Phonological Theory*, Oxford, Blackwell, 817–838.
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.
Jenkins, J., 2000, *The phonology of English as an international language*, New York, Oxford University Press.
Jespersen, O., 1904, *Lehrbuch der Phonetik*, Leipzig, Teubner.
Kelly, M.H., 2004, "Word onset patterns and lexical stress in English", *Journal of Memory and Language*, 50, 3, 231–244. DOI: https://doi.org/10.1016/j.jml.2003.12.002.
Kenworthy, J., 1987, *Teaching English pronunciation*, London, Longman.
Kim, A.R., R. Lee, J.S. Han, E.-J. Lee, 2017, "Phonological characteristics of 4- to 5-Year-Old children in nonword repetition under different phonetic complexity", *Communication Sciences & Disorders*, 22, 3, 445–457. DOI: 10.12963/csd.17419.
Kim, H.-K., 2000, *The interlanguage phonology of Korean learners of English: A computational implementation of optimality theoretic constraints*, Doctoral dissertation, Georgetown University.
Kormos, J., 2006, *Speech production and second language acquisition*, Mahwah, New Jersey, Lawrence Erlbaum Associates.
Kroll, J.F., P.E. Dussias, C.A. Bogulski, J.R.V. Kroff, 2012, "Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition", *Psychology of Learning and Motivation*, 56, 229–262.
Kubozono, H., 1999, *Nihongo no Onsei* [Japanese phonology], Tokyo, Iwanami.
Kwon, B.-Y., 2006, *Korean speakers' production of English consonant clusters: Articulatory and perceptual accounts*, Doctoral dissertation, Michigan State University.
Lado, R., 1957, *Linguistics across cultures*: *Applied linguistics for language teachers*, Ann Ardor, University of Michigan Press.
Leopold, W.F., 1970, *Speech development of a bilingual child: A linguist's record, Volume 3: Grammar and general problems in the First Two Years*, New York, AMS Press.
Mack, M., 1997, "The monolingual native speaker: Not a norm, but still a necessity", *Studies in the Linguistic Sciences*, 27, 113–146.

MacLeod, A.A.N., C. Stoel-Gammon, 2010, "Are bilinguals different? What VOT tells us about bilinguals", *Journal of Multilingual Communication Disorders*, 3, 2, 118–127.
MacNamara, John, 1967, "The linguistic independence of bilinguals", *Journal of Verbal Learning and Verbal Behavior*, 6, 5, 729–736.
Magloire, J., K. P. Green, 1999, "A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish", *Phonetica*, 56, 158–185.
Major, R., 1987, "English voiceless stop production by speakers of Brazilian Portuguese", *Journal of Phonetics*, 15, 197–202.
Maryn, Y., F. Ysenbaert, A. Zarowski, R. Vanspauwen, 2017, "Mobile communication devices, ambient noise, and acoustic voice measures", *Journal of Voice*, 31, 2, 248.e11–248.e23.
McCully, C., 2009, *The sound structure of English: An introduction*, New York, Cambridge University Press.
McLeod, S., 2010, "Laying the foundations for multilingual acquisition: An international overview of speech acquisition", in M. Cruz-Ferreira (ed.), *Multilingual norms*, Berlin, Peter Lang, 53–71.
Meisel, J., 2004, "The bilingual child", in T. Bhatia, W. Ritchie (eds), *The handbook of bilingualism*, Amsterdam, Benjamins, 91–113.
Mitchell, R., F. Myles, 2004, *Second language learning theories*, 2nd ed., London, Hodder Arnold.
Moulton, W.G., 1962, *The Sounds of English and German*, Chicago, University of Chicago Press.
Nathan, G.S., W. Anderson, B. Budsayamongkon, 1987, "On the acquisition of aspiration", in G. Ioup, S. Weinberger (eds), *Interlanguage phonology*, Rowley, MA, Newbury House, 204–212.
Nicoladis, E., 2006, "Cross-linguistic transfer in adjective-noun strings by preschool bilingual children", *Bilingualism: Language and Cognition*, 9, 1, 15–32.
Nogita, A., 2016, "Arguments that Japanese [Cj] s are complex onsets: Durations of Japanese [Cj]s and Russian [Cj]s and blocking of Japanese vowel devoicing", *Working Papers of the Linguistics Circle*, 26(1), 73–99.
Odlin, T., 1989, *Language transfer: cross-linguistic influence in language learning*, Cambridge, Cambridge University Press.
Ohata, K., 2004, "Phonological differences between Japanese and English: Several potentially problematic areas of pronunciation for Japanese ESL/EFL learners", *Asian EFL Journal*, 6, 1–19. http://www.asian-efl-journal.com/Dec_04_ko.pdf.
Oliveira, G., G. Fava, M. Baglione, M. Pimpinella, 2017, "Mobile digital recording: Adequacy of the iRig and iOS device for acoustic and perceptual analysis of normal voice", *Journal of Voice*, 31, 2, 236–242.
O'Neal, G., 2015, "Consonant clusters and intelligibility in English as a Lingua Franca in Japan: Phonological modifications to restore intelligibility in ELF", *Pragmatics and Society*, 6, 4, 615–636.
Ostapenko, O., 2005, "The optimal L2 Russian syllable onset", in *Linguistics students organization working papers in linguistics 5: Proceedings of the workshop in general linguistics*, University of Wisconsin-Madison, 140–151.
Paradis, J., F. Genesee, 1996, "Syntactic acquisition in bilingual children: Autonomous or interdependent?", *Studies in Second Language Acquisition*, 18, 1, 1–25.
Parker, S., 2011, "Sonority", in M. van Oostendorp, C. J. Ewen, E. Hume, K. Rice (eds), *The Blackwell companion to phonology*, volume 2, Hoboken, New Jersey, Wiley-Blackwell, 1160–1184.
Peterson, G.E., I. Lehiste, 1960, "Duration of Syllable Nuclei in English", *Journal of the Acoustical Society of America*, 32, 6, 693–703.
Rauber, A.S., B.O. Baptista, 2004, "The production of English initial /s/ clusters by Portuguese and Spanish EFL speakers", *Revista de Estudos da Linguagem*, 12, 2, 459–473. DOI:10.17851/2237-2083.12.2.459-473.
Riney, T., J. Anderson-Hsieh, 1993, "Japanese pronunciation of English", *JALT Journal*, 15, 1, 21–36.
Saussure, F. de, 1916, *Cours de linguistique générale*, Paris, Payot.

Scherling, J., 2012, *Japanizing English: Anglicisms and their impact on Japanese*, TŸbingen, Narr Verlag.
Schreier, D., 2005, *Consonant change in English worldwide: Synchrony meets diachrony*, New York, Palgrave Macmillan.
Selkirk, E., 1984, *Phonology and syntax: The relation between sound and structure*, Cambridge, Mass., MIT Press.
Serratrice, L., 2013, "Cross-linguistic influence in bilingual development: Determinants and mechanisms", *Linguistic Approaches to Bilingualism*, 3, 1, 3–25.
Shockey, L., 2003, *Sound patterns of spoken English*, Malden MA, Blackwell.
Sievers, E., 1881, *Grundzüge der Phonetik*, Leipzig, Breitkopf und Hartel.
Simonet, M., 2014, "Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals", *Journal of Phonetics*, 43, 26–37.
Smith, D.L., 2006, "English in disguise: Japanese renditions of words lifted from English", *Asian Englishes*, 9, 2, 68–79.
Snow, C.E., K. Hakuta, 1992, "The costs of monolingualism", in J. Crawford (ed.), *Language loyalties: A source book on the official English controversy*, Chicago, The University of Chicago Press, 384–394.
Steriade, D., 1982, *Greek prosodies and the nature of syllabification*, Doctoral dissertation, MIT.
Stockwell, R.P., J.D. Bowen, 1965, *The sounds of English and Spanish*, Chicago, The University of Chicago Press.
Takeshita, Y., 2010, "East Asian Englishes: Japan and Korea", in Andy Kirkpatrick (ed), *The Routledge handbook of world Englishes*, New York, Routledge, 265–281.
Thompson, I., 1987. "Japanese Speakers" In Michael Swan, Bernard Smith (eds), *Learner English*, Cambridge, Cambridge University Press, 212–223.
Uloza, V., N. Ulozaitė-Stanienė, T. Petrauskas, R. Kregždytė, 2021, "Accuracy of acoustic voice quality index captured with a smartphone – Measurements with added ambient noise", *Journal of Voice*, 1–8. DOI: 10.1016/j.jvoice.2021.01.025.
Vance, T.J., 1987, *An introduction to Japanese phonology*, Albany, NY, State University of New York Press.
Van der Woerd, B., M. Wu, V. Parsa, P.C. Doyle, K. Fung, 2020, "Evaluation of acoustic analyses of voice in nonoptimized conditions", *Journal of Speech, Language, and Hearing Research*, 63, 12, 3991–3999.
Van Hell, J. G., T. Dijkstra, 2002, "Foreign language knowledge can influence native language performance in exclusively native contexts", *Psychonomic Bulletin & Review*, 9, 780–789.
Von Studnitz, R. E., D.W. Green, 2002, "Interlingual homograph interference in German-English bilinguals: Its modulation and locus of control", *Bilingualism: Language and Cognition*, 5, 1, 1–23.
Weinberger, D.R., 1987, "Implications of normal brain development for the pathogenesis of schizophrenia", *Archives of General Psychiatry*, 44, 660–669.
Weinreich, U., 1953, *Languages in contact: Findings and problems*, The Hague, Mouton.
White, L., S.L. Mattys, 2007, "Calibrating rhythm: First language and second language studies", *Journal of Phonetics*, 35, 501–522.
Wu, Q., 2017, "An iPhone-based binaural recorder for sound quality analysis", *Sound & Vibration*, 51, 4, 16–17.
Wu, Y., G. Thierry, 2010a, "Chinese-English bilinguals reading English hear Chinese", *Journal of Neuroscience*, 30, 22, 7646–7651.
Wu, Y., G. Thierry, 2010b, "Investigating bilingual processing: the neglected role of language processing contexts", *Frontiers in Psychology* 1, 1–6. https://www.frontiersin.org/articles/10.3389/fpsyg.2010.00178/full.
Yavaş, M., M. Someillan, 2006, "Acquisition of /s/-clusters in Spanish–English bilinguals", *Journal of Multilingual Communication Disorders*, 3, 1, 50–55.
Zhang, C., J. van de Weijer, J. Cui, 2006, "Intra- and inter-speaker variations of formant pattern for lateral syllables in Standard Chinese", *Forensic Science International*, 158, 117–124.