

Production of coronal stops by simultaneous bilingual adults*

MEGHA SUNDARA
LINDA POLKA
SHARI BAUM
McGill University

This study investigated acoustic-phonetics of coronal stop production by adult simultaneous bilingual and monolingual speakers of Canadian English (CE) and Canadian French (CF). Differences in the phonetics of CF and CE include voicing and place of articulation distinctions. CE has a two-way voicing distinction (in syllable initial position) contrasting short- and long-lag VOT; coronal stops in CE are described as alveolar. CF also has a two-way voicing distinction, but contrasting lead and short-lag VOT; coronal stops in CF are described as dental. Acoustic analyses of stop consonants for both VOT and dental/alveolar place of articulation are reported. Results indicate that simultaneous bilingual as well as monolingual adults produce language-specific differences, albeit not in the same way, across CF and CE for voicing and place. Similarities and differences between simultaneous bilingual and monolingual adults are discussed to address phonological organization in simultaneous bilingual adults.

Introduction

In recent years there has been increased interest in the speech produced by children exposed to two languages from birth. A majority of the studies on the developing phonology of children exposed to two languages rely on transcription data to characterize speech production behavior in the two languages of the simultaneous bilingual child (Leopold, 1949/1970; Vogel, 1975; Celce-Murcia, 1978; Schnitzer and Krasinski, 1994, 1996; Johnson and Lancaster, 1998).

Although transcription data provide a summary of the entire phonological system of the individual being tested, getting a fine-grained account of language-specific characteristics from transcription studies is problematic. As Macken and Barton (1980) point out, sub-phonemic differences between two physically (acoustically or articulatorily) non-identical tokens may not be detected by the mature phonological system of the adult listener (e.g. /r/-/l/ differences to a Japanese transcriber) for numerous reasons. Simultaneous bilingual children may use secondary cues that are appropriate, but not adequate in the absence of the primary cue to signal contrasts. They may use cues inappropriate for the speech community or be more or less variable in the underlying gesture, or temporal or spectral acoustic cues – differences that may remain imperceptible. Thus, differences between

segments produced by simultaneous bilinguals are likely to be underestimated due to the limitations of the transcriber's perceptual system.

Instrumental analyses provide a consistent and reliable approach to measuring language-specific implementation in the two languages of bilinguals. Previous research using acoustic analyses of productions by children exposed to two languages from birth has focused on differences in the implementation of voicing across the two languages of the bilingual as measured by voice onset time (VOT; Watson, 1990; Khattab, 2000; Johnson and Wilson, 2002). VOT is a significant acoustic correlate of the voiced–voiceless distinction across the world's languages, and probably the most widely reported measure. In their seminal cross-language investigation, Lisker and Abramson (1964) describe VOT as the time interval between the burst that signals the release of the stop closure and the onset of periodicity indicative of laryngeal vibration. Typically, VOT values less than 0 ms are referred to as lead VOT, values between 0 and 30 ms as short-lag VOT and values greater than 30 ms as long-lag VOT.

Johnson and Wilson (2002) present developmental VOT production data from two sisters, 4 years and 8 months old and 2 years and 10 months old, both exposed to Japanese and English from birth. Japanese and English, both have a two-way distinction between voiced and voiceless stops at all places of articulation. However, the VOT pattern signaling the voicing contrast is different in the two languages. English speakers typically produce voiced stops with short-lag (or sometimes lead) VOT whereas they produce voiceless stops with long-lag VOT. In Japanese, voiced stops are typically produced with lead VOT (or sometimes short-lag VOT) whereas voiceless stops are produced with either short- or long-lag VOT (Homma, 1980). Because both girls were aware that their

* This is part of M. Sundara's doctoral thesis. It was supported by an Internal Social Sciences and Humanities Research Council grant from McGill University to M. Sundara (202686) and a Discovery Grant to L. Polka from Canada's Natural Science and Engineering Research Council (RGPIN 105397). We would like to thank Georgina Hernandez for her help with collecting the French samples and Fred Genesee, Susan Guion and two anonymous reviewers for comments on previous versions of this paper.

Address for correspondence

Megha Sundara, Institute for Learning & Brain Sciences, Box 357988, University of Washington, Seattle, WA 98195-7988, USA

E-mail: msundara@u.washington.edu

mother (Japanese) and father (Canadian English) spoke different languages, the authors elicited tokens in the two languages by asking the children what their mother/father would call pictured objects. Sessions were recorded in the presence of both parents. Johnson and Wilson (2002) report that only the 4-year-old produced voiceless stops with different VOT values in the two languages. Neither of the two sisters differentiated between voiced stops in Japanese and English; in both languages /d/ was produced with short-lag VOT. Johnson and Wilson also indicate that both sisters sounded native-like in the two languages. Note that, in this case, transcription data would have obscured differences in VOT evident in the acoustic analyses.

Johnson and Wilson's study has several limitations. As mentioned previously, neither of the two bilingual children with simultaneous exposure in Japanese and English studied by Johnson and Wilson produced voiced stops with lead VOT. Because monolingual children have also been reported to have difficulty producing lead VOT (Zlatin and Koenigsknecht, 1976; Macken and Barton, 1980), the absence of lead VOT in voiced stops produced by bilingual children may not be a consequence of their exposure to two languages. It has been suggested that production of lead VOT requires motorically complex timing of laryngeal and oral gestures (Kent, 1992) and, consequently, is difficult for children with immature sensory and motor systems. In the absence of data from monolingual children using a comparable methodology, it is not possible to distinguish between the two accounts. Furthermore, Johnson and Wilson recorded the children in the presence of both parents, encouraging a bilingual 'mode'. There is substantial evidence that language mode influences speech production and perception by simultaneous bilinguals and L2 learners (Lanza, 1992; Genesee, Boivin and Nicoladis, 1996; Grosjean, 1997; Treffers-Daller, 1997). Specifically, recording bilingual speakers in bilingual settings is thought to activate both languages simultaneously; as a consequence, language-specific differences between the two languages are likely to be underestimated.

Research in Khattab (2000), comparing VOT data from older children exposed to Arabic and English from birth, and their monolingual peers, presents evidence to disambiguate differences relating to developmental immaturity and dual language exposure. Like Japanese, Arabic also has a two-way voicing distinction. In Arabic, voiced stops are produced with lead VOT whereas voiceless stops are produced sometimes with short-lag and sometimes with long-lag VOT (Yeni-Komshian, Caramazza and Preston, 1977). Khattab (2000) analyzed stop consonants produced by 3 bilingual children aged 5, 7 and 10 and their monolingual peers. Bilingual children were asked to name pictures in two sessions, one in each language with different interlocutors. Khattab (2000) reports that the children with simultaneous exposure

produced voiceless stops in the two languages like their monolingual peers – in English with long-lag VOT and in Arabic, (with the exception of the 5-year-old) with short-lag VOT. Similar to the results reported by Johnson and Wilson (2002), the monolingual Arabic children but not the bilingual children produced voiced tokens in Arabic with lead VOT. The bilingual children produced English-like, short-lag VOT for voiced tokens in both languages. However, all the children, except the 5-year-old, produced significantly different values of VOT for voiced stops in English and Arabic. Because monolingual children of the same age were able to produce lead VOT, the inability of bilingual children to produce lead VOT is not likely to be due to the developmental immaturity of their sensory-motor systems. It is more likely to be a consequence of the language-experience of these bilingual children – either due to the more limited (or more varied) exposure to Arabic received by bilingual children, the relative difficulty of producing lead VOT, or a combination of these factors.

Because both Johnson and Wilson (2002) and Khattab (2000) report data from one subject at each age for different language pairs, their findings have limited generalizability. In a larger study, Watson (1990) presents VOT data from five children per group, aged 6, 8 and 10, exposed to both British English and European French from birth, their monolingual peers and monolingual adults. European French has a two-way voicing distinction; voiced stops are produced with lead VOT whereas voiceless stops are produced with short-lag VOT (Caramazza and Yeni-Komshian, 1974; Watson, 1990). The bilingual children tested by Watson were resident in Paris, had at least one English-speaking parent, were being educated bilingually and were chosen with the help of teachers as having native-like performance in both languages. Children were asked to repeat words presented by a different interlocutor in each language. Similar to Johnson and Wilson (2002) and Khattab (2000), Watson also reports that bilingual children at all three ages produced voiceless stops in English and French with different VOT values, but they produced voiced stops in both languages with similar, short-lag VOT. However, unlike data reported in Khattab (2000), Watson's bilingual children produced VOT values comparable to their monolingual peers at all ages. Thus, the group data from Watson's study provides no evidence for a delay in the ability of bilingual children to produce appropriate VOT values for voiced and voiceless stops in their two languages.

Because voicing can be signaled by cues other than VOT (Summerfield and Haggard, 1977; Lisker, 1978; Ohde, 1984), Watson also presents data on onset of fundamental frequency, onset of first formant frequency and length of following vowel for all groups. Children with simultaneous exposure produced greater differences in F1 onset frequency in French and began to produce the vowel

length distinction that signals voicing difference earlier than their monolingual peers. It is important to note that although the bilingual children in Watson's study differed in phonetic implementation from their monolingual peers, adult listeners in both languages perceived bilingual children as being native-like. Thus, Watson (1990, 1996) suggests that the speech production of bilingual children may be qualitatively different from their monolingual peers while remaining perceptually undetected by native listeners.

Results from Watson's study underscore the need for measuring more than one acoustic correlate of phonetic/phonological distinctions of interest. The differences between lead VOT produced by monolingual and simultaneous bilingual children reported by Khattab (2000) and F1 onset and vowel length reported by Watson (1990) are consistent with differences in rate of development between bilinguals and monolinguals. However, Watson's report of differences in F1 onset and vowel length, but not VOT, point to qualitative or pattern differences between bilingual and monolingual children. In the absence of data from adult simultaneous bilinguals, it is not possible to determine whether differences between children learning one and two languages are best described as differences in rate or as pattern of development.

To date, Guion (2003) presents the only acoustic data on speech production by simultaneous bilingual adults. Her study of Quichua- and Spanish-speaking adults suggests that differences between monolingual and simultaneous bilingual children are related to differences in rate of development. Guion (2003) compared F1 and F2 for vowels produced by speakers exposed to both Quichua and Spanish from birth, Spanish monolinguals, and early, mid and late L2 learners of Spanish. The late L2 learners were native Quichua speakers exposed to Spanish between 15 and 25 years of age who had not learned to produce any vowel categories in Spanish. This group served as a native Quichua control because monolingual Quichua speakers were not available. Ecuadorian Quichua has a three-vowel inventory, /i a u/, whereas Spanish has a five-vowel inventory, /i e a o u/. The early L2 learners (age of acquisition between 5 and 7 years) were not significantly different from monolingual Spanish speakers in their production of Spanish vowels; however, they were significantly different from native Quichua speakers in their production of Quichua vowels. In contrast, simultaneous bilinguals produced vowels in Quichua comparable to native Quichua speakers and produced vowels in Spanish comparable to Spanish monolinguals. Guion's (2003) data suggest that only adults exposed to two languages from birth produce vowel categories similar to both monolingual groups.

However, because all the developmental data on bilingual children is on VOT production, there is a need for empirical data on VOT from simultaneous bilingual

adults. There is also a need for further research comparing monolingual and bilingual production of other consonant features and segments. Such data will help determine whether differences observed between monolingual and bilingual children relate to differences in rate or pattern of acquisition. If simultaneous input in two languages only alters the rate of acquisition, differences between bilingual and monolingual production should be observed only during development and simultaneous bilingual adults should resemble monolingual adults. On the other hand, if differences between monolingual and bilingual children are indicative of differences in pattern of development, differences between the two groups can be expected to persist into adulthood.

Data from simultaneous bilingual adults are also pertinent to our understanding of the extensive literature on speech production behavior of second language (L2) learners exposed sequentially to two languages (Caramazza, Yeni-Komshian, Zurif and Carbone, 1973; Flege and Eefting, 1987a, b; Mack, 1989; Hazan and Boulakia, 1993; Grosjean and Miller, 1994; Flege, Munro and MacKay, 1995; Flege, Takagi and Mann, 1995; Flege, Schirru and MacKay, 2003). Studies of L2 learners typically involve two questions: Do L2 learners systematically distinguish between productions in L1 and L2? Are they different from native or monolingual controls?

This is illustrated in the classic study by Caramazza et al. (1973). They compared VOT implementation in Canadian English (CE) and Canadian French (CF) by monolingual CE and CF speakers and native speakers of CF who had begun to acquire CE before 7 years of age. Caramazza et al. report that for isolated words in syllable initial position, both CE and CF have two-way voicing distinctions for stops. However, the VOT pattern signaling the voicing contrast is different in the two languages. Monolingual CE speakers, like American English speakers, typically produce voiced stops with short-lag (or sometimes lead) VOT whereas they produce voiceless stops with long-lag VOT. Monolingual CF speakers produce voiced stops with lead VOT or short-lag VOT, whereas they produce voiceless stops with a short-lag VOT.

Caramazza et al. (1973) report that early L2 learners of CE produced significantly different VOT for voiceless stops in CE and CF, but were still significantly different from monolingual speakers. They produced /t/ in French (28 ms) like monolingual CF speakers (23 ms); however, they produced /t/ in English (48 ms) with compromise or intermediate VOT values, falling between the VOT for voiceless stops by monolingual CE (70 ms) and CF speakers (23 ms). Caramazza et al. do not report means or statistical comparisons of VOT for voiced stops. However, from the frequency distribution of VOT values presented in their paper, it appears that early L2 learners

produce similar VOT values for /d/ in CE and CF. In both languages, voiced tokens are produced with lead VOT or short-lag VOT. Thus, Caramazza et al.'s data provide evidence for language-specific production of VOT by early L2 learners for /t/, but not /d/, that is different from the VOT values produced by monolinguals.

Based on differences between production by L2 learners and L1 controls, several authors have concluded that L2 learners are unable to form distinct systems for L1 and L2. Instead, mutual influences of L1 and L2 determine production and perception of sounds in the two languages of L2 learners (for reviews see Best, 1995; Flege, 1995; Guion, 2003). However, interactions that cause L2 learners to differ from monolinguals could result from the later age of acquisition of the second language in L2 learners or simply be due to early exposure to two languages. If mutual influences, i.e. interactions between the two languages of L2 learners are due to the later age of acquisition of the second language, simultaneous bilingual adults are not expected to be different from monolinguals. To address this question, in the current study, speech production by simultaneous bilingual adults was compared with production by monolinguals.

The present study

The present study was designed to investigate speech production by adult simultaneous bilinguals. For this purpose, coronal stops produced by adults who have been exposed to CE and CF from birth were analyzed acoustically. CE and CF coronal stops differ in voicing as well as place of articulation. Voicing differences in CE and CF can be captured by differences in VOT as shown by Caramazza et al. (1973) and replicated by Sundara (2005).

Like American English, in initial position, CE coronal stops are phonetically transcribed as alveolar whereas CF coronal stops are transcribed as dental (Picard, 1987, 2001). Currently, there are no articulatory data confirming that CE tokens are alveolar and CF tokens dental. However, for ease of description, coronal stops in the two languages will be referred to by their phonetic descriptions.

Acoustic differences between coronal stops thought to differ in dental and alveolar place have been previously analyzed using relative burst intensity and burst spectral measures as indexed by mean, standard deviation, skewness and kurtosis of burst frequency (Jongman, Blumstein and Lahiri, 1985; Stoel-Gammon, Williams and Buder, 1994). Compared to CF stops, bursts associated with CE stops are louder, have a higher burst frequency and are more compact as measured by standard deviation, with a more peaked spectral shape as measured by the kurtosis of burst frequency (Sundara, 2005).

In this study, simultaneous bilingual and monolingual¹ adults were recorded while producing /d/- and /t/-initial words in sentences. Coronal stops produced by monolinguals and bilinguals were analyzed for differences in VOT, relative burst intensity and burst spectral measures. Data from monolingual subjects were included to establish acoustic characteristics of voicing and place differences between coronal stops in CE and CF, and also to allow for direct comparisons between productions by monolinguals and bilinguals. Data from bilingual subjects were used to address two specific questions. First, do bilinguals produce language-specific differences in voicing and place in CE and CF? Second, how (if at all) do bilinguals differ from the monolingual groups? This is the first set of data on acoustic implementation of voicing and place differences in adults who have been exposed to two languages simultaneously from birth.

Method

Subjects

Seventeen adults (mean age = 24; range = 22–35) were recorded. Six were monolingual speakers of CE (3 male & 3 female); six were monolingual speakers of CF (3 male & 3 female); and five (2 male and 3 female) were simultaneous bilingual CE/CF speakers. The subjects had no history of speech, language, or hearing impairment. Each subject's language background was assessed using a detailed language questionnaire, including self-ratings of language ability using a scale from 1 to 7, where 7 represents native-like ability and 1 represents no ability. This was followed by an interview with a proficient bilingual research assistant. A language sample describing a picture story ("Frog, where are you?" by Mayer, 1969) was collected from each subject. These samples were presented to three monolingual speakers of CF and CE. They were asked to rate the sample on a scale from 1 to 7, where 7 represents native-like ability and 1 represents no ability.

To be included in the monolingual CE (or CF) group, subjects had to meet the following five criteria. First, subject's parents were monolingual speakers of CE (or CF). Second, subjects were schooled only in CE (or CF). Third, they rated their ability in their native language with a minimum of 6 on a scale of 1–7 (where 7 represents native-like ability and 1 represents no ability at all). If they had any knowledge of the non-native language, they rated it below 3 on the same scale. This was necessary

¹ Although most people educated in Canada receive some formal instruction in both languages at school, this instruction is mainly in reading and writing with minimal emphasis on speaking or listening skills. The term monolingual is used here to describe native speakers with little or no proficiency in any other language.

Table 1. *Test words in Canadian English and Canadian French. Only word initial /d/ and /t/s were analyzed*

| Canadian English | | Canadian French | |
|------------------|---------|-----------------|----------|
| docile | toffee | docile | toffee |
| doctor | topic | docteur | topique |
| dopey | total | doper | total |
| dodo | topaz | dodo | topaz |
| deadly | textile | detter | textile |
| despot | texture | despote | texture |
| dagger | tablet | dadais | tablette |
| dapper | taxi | datcha | taxi |

as most people in Canada receive some education in the other language. The bilingual interviewer confirmed their proficiency in the non-native language during the interview. Fourth, they had spent no time in a foreign country where a language other than their native language was spoken. Fifth, native CE (or CF) listeners rated their speech sample with a minimum of 6 on a scale from 1 to 7.

To be included in the simultaneous bilingual group, subjects had to meet the following five criteria. First, subjects had learned both CE and CF simultaneously at home from their parents, each of whom was a native speaker of one language and communicated with the subject primarily in that language. Second, their schooling was completed either in bilingual schools, or at different points in CE and CF. Third, subjects rated their knowledge of both languages with a minimum of 6 on a scale of 1–7. A fluently bilingual interviewer confirmed their self-ratings of level of proficiency in both languages during the interview. Fourth, they lived in bilingual areas like Montreal or the Ottawa region and were using both languages consistently within the home and the work context. Fifth, native CE and CF listeners rated their speech samples in both languages with a minimum of 6 on a scale from 1 to 7.

Six additional monolingual subjects (2 CE male and 1 CE female; 2 CF female and 1 CF male) were recorded but excluded from the analyses because native listeners rated their speech samples lower than 6. Five additional bilingual subjects (3 male and 2 female) were also recorded, but excluded from analyses because native listeners rated their speech samples in one (or both) language lower than 6.

Stimuli

All subjects were recorded producing target bisyllabic real words (Table 1) with coronal stops in word initial position in a soundproof booth using a tabletop AKG C1000S microphone and a Tascam DA-30 digital audio recorder. Subjects read target words, twice produced in

sentences, followed by twice in isolation. The English words were produced in the carrier phrase “Now I say – again”; the French words were produced in the carrier phrase “Maintenant je dis – encore.” Only the tokens produced in sentence context are reported here. Different interlocutors recorded bilinguals in two separate sessions, one in CE and one in CF. The first author conducted the English sessions; all subjects were aware that she was a native speaker of Indian English and her communication skills in CF were very limited. A Spanish and French bilingual research assistant conducted the French sessions. Although the research assistant did speak English, having learned it at school, her English was accented and subjects were aware that she was not very comfortable conversing in English and clearly preferred French. The sessions and testers were kept separate to ensure that bilinguals were in a (relatively) monolingual mode. The order of recording in the two languages was counterbalanced across subjects. Acoustic data from isolated word productions by these monolingual speakers are reported in Sundara (2005).

Phonologically similar words were selected from the two languages to minimize differences in the target coronal stop due to differences in the phonetic context in which it was produced in the two languages. Care was taken to ensure that the consonant following the target syllable was a fricative, affricate or stop, so that the syllable boundary could be easily identified on the spectrographic and waveform display. Additionally, articulatory descriptions of vowels following coronal stops in the target words (/ɛ/, /æ/, /o/ and /ɔ/) were identical across the two languages (Picard, 1987, 2001). Vowel formants were measured at midpoint to confirm overlap in acoustic space for CE and CF tokens. Because few monosyllabic words met the above-mentioned criteria, bisyllabic words were used despite differences in phrasal accent allocation in CE and CF. Although there is little research on the effect of phrasal accent on burst intensity and spectral measures, a recent study (Cole, Choi, Kim and Hasegawa-Johnson, 2003) indicates that at least within English, burst amplitude in accented and unaccented syllables was not significantly different. Tokens were digitized at 22050 Hz and 16-bit quantization. Subsequently, the initial CV was excised from each target word. Acoustic analyses are reported for these syllables.

Tokens excluded

Monolingual as well as bilingual speakers produced /d/ tokens in CF without clear bursts. Of the monolingual speakers, one male and two female speakers produced 25, 16.7 and 45.4% of the tokens without clear bursts. Of the bilingual speakers, two male and two female speakers produced 36, 17, 7 and 20% of the tokens without clear bursts. Stops without clear bursts were also observed for

voiced stops produced at other places of articulation, in initial as well as medial position. Number of tokens produced without clear bursts were comparable across monolingual and bilingual speakers; they were excluded from all analyses.

Acoustic analyses

A total of 602 tokens, 322 CE (178 produced by monolinguals and 144 by bilinguals) and 280 CF tokens (150 produced by monolinguals and 130 by bilinguals) were analyzed using PRAAT (Boersma and Weenink, 2005). To analyze vowel formants, VOT, burst intensity and burst spectral properties of /d/ and /t/ in each excised CV, five cursor positions were identified using a waveform display supplemented by a wide-band spectrographic display: onset of first periodic pattern before the burst (if any), onset of the burst, offset of the burst, onset of first periodic pattern after the burst signaling vowel onset and vowel offset.

VOT was measured as the time between the onset of the first clearly periodic pattern and the onset of the burst (Lieberman and Blumstein, 1988). Only tokens where the glottal vibration was continuous with the burst were considered to have lead VOT. This was done in order to distinguish between lead VOT and the glottal vibration associated with the preceding vowel.

Burst intensity and shape of the burst spectrum were calculated over the entire burst beginning at consonantal release. Size of the analysis window thus varied from token to token; it was determined by the duration of burst. When calculating burst intensity measures for voiceless aspirated stops in CE, aspiration was not included in the analysis window. To adjust for differences in overall intensity across speakers, relative burst intensity (RI) was calculated by subtracting the intensity of the burst (in dB) from the maximum intensity of the following vowel (in dB). On this measure, a softer burst is expected to have a greater intensity difference from the subsequent vowel.

Shape of the burst spectrum as characterized by the 4 spectral moments – mean, standard deviation, skewness and kurtosis – was measured (Forrest, Weismer, Milenkovic and Dougall, 1988; Stoel-Gammon et al., 1994). The four spectral moments index four independent features of the energy distribution over frequency to derive average energy concentration (mean frequency) and spectral shape as indexed by spread of frequency around the mean (standard deviation), the symmetry or tilt of the distribution (skewness) and the degree of its peakedness (kurtosis). Mean, standard deviation, skewness and kurtosis of burst frequency were derived from the power spectra over the entire burst duration for frequencies up to 11025 Hz. Bursts were pre-emphasized prior to making spectral measurements. Voiced tokens in CF, and sometimes in CE, are produced with prevoicing.

Prevoicing is characterized by regular low frequency glottal vibration during stop closure and sometimes through the burst. To compare intensity and spectral measures for voiced and voiceless stop consonants, all stops with lead VOT were filtered using a 200 Hz high-pass filter to remove the effects of voicing (a similar technique was used by Jongman et al., 1985).

Formant frequencies (F1 and F2) are useful for distinguishing vowels acoustically. Thus, F1 and F2 were measured at mid-point between vowel onset and offset to confirm that the vowels in CE and CF overlapped acoustically.² For each vowel, monolingual and bilingual speakers of both genders showed considerable overlap in the formant frequencies of CE and CF tokens. Compared to the vowel space in CE, the CF vowel space was shifted upward in F2 and downward in F1 for both male and female speakers.

Results

Results from the analyses are presented in three sections. First, a comparison of monolingual CE and CF speakers' productions is presented. Then, a comparison of bilingual speakers' productions in CE and CF are presented; analyses are reported to determine whether or not bilingual speakers produce language-specific differences in the two languages. In both sections, results for VOT are summarized, followed by results on burst amplitude and spectral measures.³ For analyses, VOT data were pooled across gender because there was no reason to expect gender differences in production of VOT. Data on each burst measure were analyzed separately. Because the main question in this paper concerns differences in production characteristics of tokens across CE and CF and not gender differences per se, data were pooled across gender for burst measures as well. Collapsing data across gender yields a more robust analysis of differences between languages.⁴ Finally, bilingual speakers' production in each language is compared with the respective monolingual groups. In this section, t-tests are presented to determine whether or not bilingual speakers behave like one or the other monolingual group.

² F1 and F2 were derived from LPC analysis with a 15 ms hamming window centered at vowel steady state. Data on vowel formants are available on request.

³ Note that although statistical significance can help identify possible cues to voicing and place differences, listeners may not perceive all statistically significant differences. Nevertheless, acoustic data are useful to narrow down the perceptually relevant cues signaling coronal place, as differences between groups that are not statistically significant are not likely to signal perceptual differences reliably.

⁴ All the analyses were also carried out for male and female speakers separately. The pattern of results for Language effects for each of the two genders is identical to that reported here. When data from male and female speakers are analyzed separately, simple effects of voicing are sometimes different from those reported here.

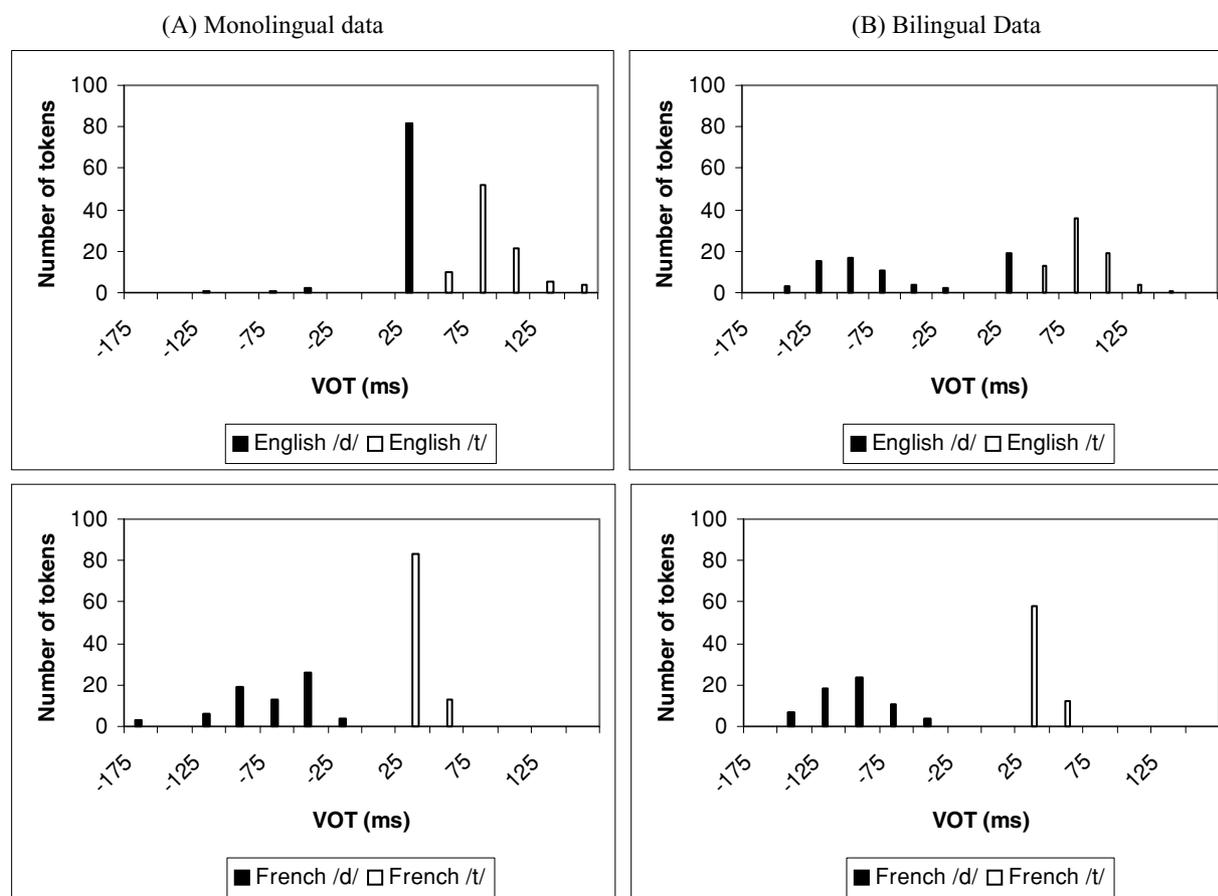


Figure 1. Distribution of VOT (ms) for /d/ (light bars) and /t/ tokens (dark bars) in Canadian English and Canadian French – (A) Monolingual data, (B) Bilingual data. Each bar represents the number of tokens produced with VOT values within a 50 ms bin.

Monolingual speakers: Between-subjects comparison

In this section, data from monolingual speakers are analyzed to confirm that they produce language-specific differences in VOT, burst intensity and spectral measures while producing coronal stops in CE and CF sentences. For this purpose, General Linear Model (GLM) repeated measures analysis of variance (ANOVA) with Language (CE and CF) as the between-subjects variable and Voicing (voiced and voiceless) as the within-subjects variable were conducted. A GLM analysis provides more power for comparisons of groups with unequal cell size. When significant, interactions were explored for simple effects of Language and Voicing using t-tests with Bonferroni's correction for multiple comparisons ($p < 0.01$).

VOT

The distributions of VOT for /d/ and /t/ in CE and CF are summarized in Figure 1A. Overall, the distribution of VOT in CE and CF was different. Monolingual CE

speakers produced 95.3% /d/ tokens with short-lag VOT (0–30 ms) and 4.7% of tokens with lead VOT (less than 0 ms); 100% of /t/ tokens were produced with long-lag VOT (greater than 30 ms). Monolingual CF speakers produced 100% of /d/ tokens with lead VOT; and 100% of /t/ tokens were produced with short-lag VOT. The differences were confirmed by the ANOVA results. The main effects of Language ($F(1, 151) = 706, p < 0.01$) and Voicing ($F(1, 151) = 1153, p < 0.01$) and the interaction of Language and Voicing ($F(1, 151) = 84, p < 0.01$) were significant. Bonferroni's post-hoc tests confirmed that Language differences were significant for /d/ and /t/; Voicing differences were also significant in both languages – /d/ tokens had a lower VOT than /t/ tokens.

Relative burst intensity

The average relative burst intensities for monolingual CE and CF speakers are summarized in Figure 2A. Recall that a smaller relative intensity is a consequence of a louder burst. Overall, monolingual CE speakers

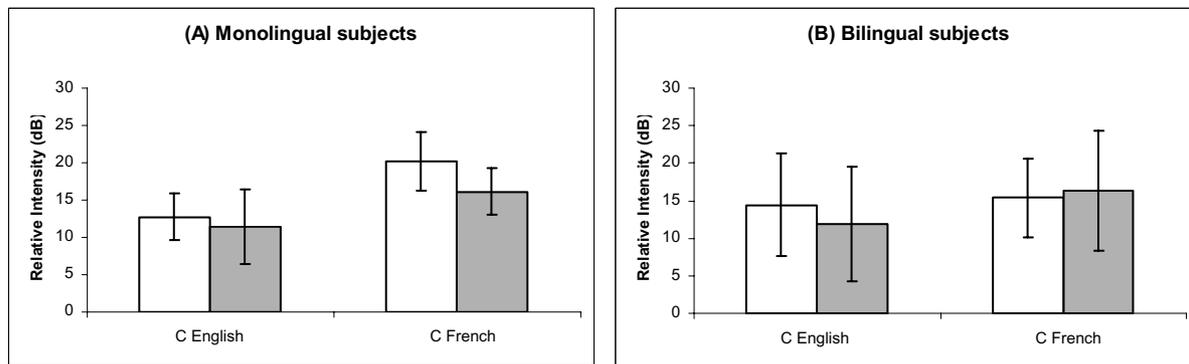


Figure 2. Average Relative Intensity (dB) for /d/ (light bars) and /t/ tokens (dark bars) in Canadian English and Canadian French – (A) Monolingual data, (B) Bilingual data. Error bars represent standard deviations.

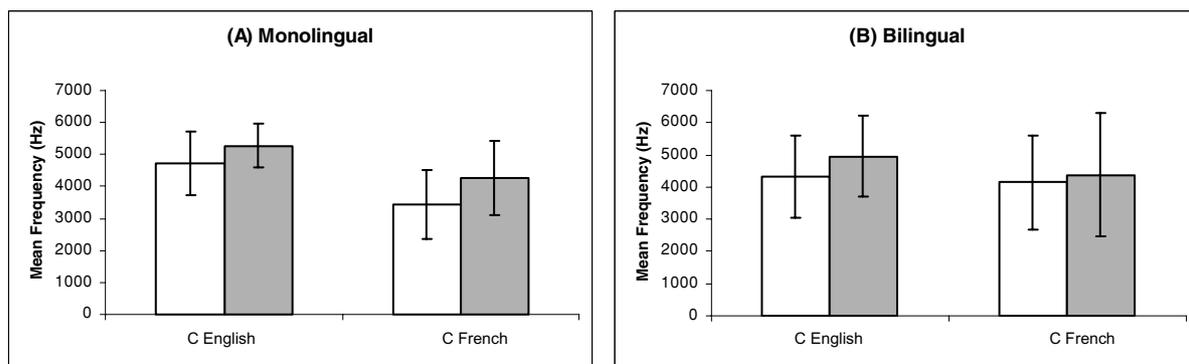


Figure 3. Average Mean Burst Frequency (Hz) for /d/ (light bars) and /t/ tokens (dark bars) in Canadian English and Canadian French – (A) Monolingual data, (B) Bilingual data. Error bars represent standard deviations.

produced louder bursts than monolingual CF speakers. In the ANOVA, the main effects of Language ($F(1, 142) = 123.8, p < 0.01$) and Voicing ($F(1, 142) = 56.2, p < 0.01$) and the interaction of Language and Voicing ($F(1, 142) = 10.6, p < 0.01$) were significant. Bonferroni's post-hoc tests confirmed that Language differences were significant for /d/ and /t/; Voicing differences were also significant in both languages – /d/ tokens had a greater relative intensity than /t/ tokens.

Mean burst frequency

The average mean burst frequencies for monolingual CE and CF speakers are summarized in Figure 3A. Overall, monolingual CE speakers produced bursts with higher mean frequency than monolingual CF speakers. In the ANOVA, the main effects of Language ($F(1, 142) = 76, p < 0.01$) and Voicing ($F(1, 142) = 44.9, p < 0.01$) were significant. Bonferroni's post-hoc tests confirmed that Language differences were significant for /d/ and /t/; Voicing differences were also significant in both languages – /d/ tokens had a lower mean frequency than /t/ tokens.

SD of burst frequency

The average SD of burst frequency for monolingual CE and CF speakers are summarized in Figure 4A. Overall, monolingual CE speakers produced bursts with smaller SD of burst frequency than monolingual CF speakers indicating that compared to CF coronals, spectra of CE coronals were compact with energy spread over a smaller range of frequencies. In the ANOVA, the main effects of Language ($F(1, 142) = 187.5, p < 0.01$) and Voicing ($F(1, 142) = 12.7, p < 0.01$) and the interaction of Language and Voicing ($F(1, 142) = 38.6, p < 0.01$) were significant. Bonferroni's post-hoc tests confirmed that Language differences were significant for /d/ and /t/; voicing differences were significant only in CE, /d/ tokens had a greater SD than /t/ tokens.

Skewness of burst frequency

The average Skewness of burst frequency for monolingual CE and CF speakers are summarized in Figure 5A. Overall, there was no consistent pattern for skewness. As observed in Figure 5A, the direction of language difference was not consistent across /d/ and /t/ tokens.

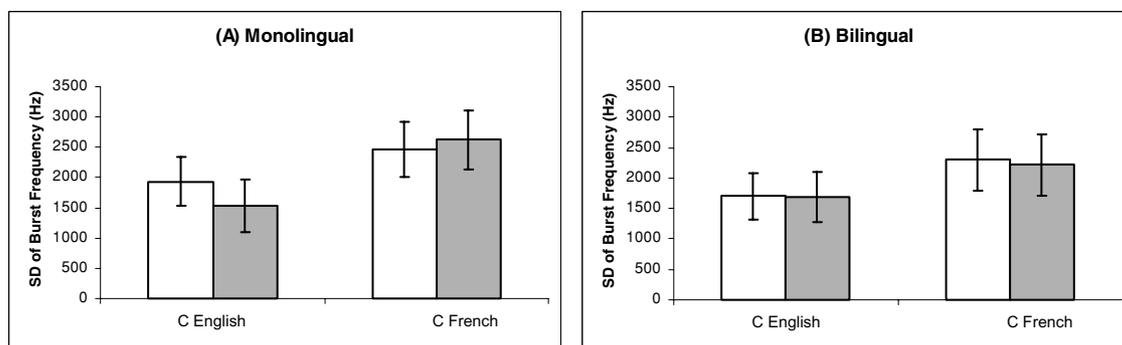


Figure 4. Average SD of Burst Frequency (Hz) for /d/ (light bars) and /t/ tokens (dark bars) in Canadian English and Canadian French – (A) Monolingual data, (B) Bilingual data. Error bars represent standard deviations.

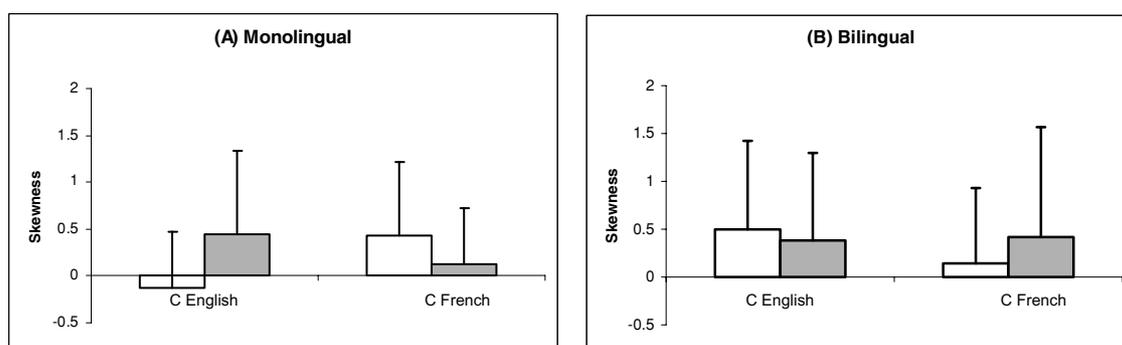


Figure 5. Average skewness of burst frequency (Hz) for /d/ (light bars) and /t/ tokens (dark bars) in Canadian English and Canadian French – (A) Monolingual data, (B) Bilingual data. Error bars represent standard deviations.

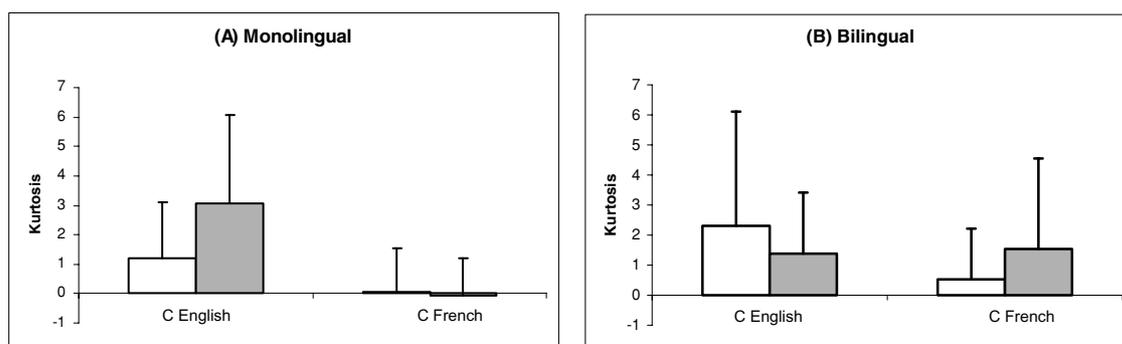


Figure 6. Average Kurtosis of burst frequency (Hz) for /d/ (light bars) and /t/ tokens (dark bars) in Canadian English and Canadian French – (A) Monolingual data, (B) Bilingual data. Error bars represent standard deviations.

Kurtosis of burst frequency

The average kurtosis of burst frequency for monolingual CE and CF speakers are summarized in Figure 6A. Overall, monolingual CE speakers produced bursts with higher kurtosis than monolingual CF speakers indicating that compared to CF coronals, spectra of CE coronals was clearly defined with well resolved peaks. In the ANOVA, main effects of Language ($F(1, 142) = 65.8, p < 0.01$) and Voicing ($F(1, 142) = 17.7, p < 0.01$) and the interaction of Language and Voicing ($F(1, 142) = 19.1, p < 0.01$)

were significant. Bonferroni's post-hoc tests confirmed that Language differences were significant for /d/ and /t/; voicing differences were significant only in CE, /d/ tokens had a lower kurtosis than /t/ tokens.

Summary

Analyses of data from monolingual speakers confirm that when tokens are produced in sentences, monolingual speakers produce language-specific differences in CE and CF. Tokens across the two languages were different

in VOT, relative intensity, mean, SD and kurtosis of burst frequency. Furthermore, within CE, /d/ tokens were different from /t/ tokens in VOT, relative intensity, mean and SD of burst frequency, whereas within CF, /d/ tokens were different from /t/ tokens in VOT, relative intensity and mean burst frequency. The pattern of results for the burst amplitude and spectral measures was similar to that reported in Sundara (2005) for bisyllabic words produced in isolation.

Bilingual speakers: Within-subject comparisons

In this section, data from bilinguals are analyzed to ascertain whether they produce language-specific differences in VOT, burst intensity and spectral measures while producing coronal stops in CE and CF. For this purpose, GLM repeated measures ANOVA with Language (CE and CF) and Voicing (voiced and voiceless) as within-subjects variables were carried out. When significant, interactions were explored for simple effects of Language and Voicing using t-tests with Bonferroni's correction for multiple comparisons ($p < 0.01$). Subsequent to analyzing group patterns, individual subject data are examined to determine how well group patterns reflect individual performance. Note that a comparison of CE and CF tokens for each subject is possible only for the bilingual group because each subject acts as his/her control.

VOT

The distributions of VOT for /d/ and /t/ in CE and CF are summarized in Figure 1B. Overall, the distribution of VOT in CE and CF was different. In CE, bilingual speakers produced 73.6% of /d/ tokens with lead VOT and 26.4% of the tokens with short-lag VOT; 100% of the /t/ tokens were produced with long-lag VOT. Thus, compared to monolingual speakers, bilingual speakers produced more /d/ tokens in CE with lead VOT. In CF, bilingual speakers produced 100% of the /d/ tokens with lead VOT; 100% of the /t/ tokens were produced with short-lag VOT. Similar to the pattern observed for monolingual speakers, there was no overlap in the distribution of VOT values for /d/ and /t/ in either CE or CF. In the ANOVA, the main effects of Language ($F(1, 132) = 162, p < 0.01$) and Voicing ($F(1, 132) = 1134, p < 0.01$) were significant. Bonferroni's post-hoc tests confirmed that Language differences were significant for /d/ and /t/; Voicing differences were also significant in both languages – /d/ tokens had a lower VOT than /t/ tokens.

Relative burst intensity

The average relative intensities for bilingual CE and CF speakers are summarized in Figure 2B. Overall, unlike the pattern observed for monolingual speakers, there was no consistent difference in relative intensity between CE and CF tokens produced by bilingual speakers. In the

ANOVA, the main effects of Language ($F(1, 46) = 12.5, p < 0.01$) and the interaction of Language and Voicing ($F(1, 46) = 10.2, p < 0.01$) were significant. Bonferroni's post-hoc tests confirmed that Language differences were significant only for /t/ tokens. As can be seen in Figure 2B, Voicing differences were in opposite directions in the two languages, and significant only in CE – /d/ tokens were softer and consequently had a greater relative intensity than /t/ tokens. Thus, similar to the pattern observed for monolingual speakers, bilingual speakers also produced relative intensity differences between /d/ and /t/ tokens in CE. However, unlike monolinguals, bilingual speakers did not produce relative intensity differences between /d/ and /t/ in CF.

A caveat is in order before interpreting null effects on this measure. Recall that in CF, /d/ tokens without clear bursts were excluded from analyses; an absent or unclear burst effectively guarantees a very large relative intensity difference between the vowel and the burst, namely a difference equal to the intensity of the vowel (about 60–65 dB). Including tokens without clear bursts in the analyses would most likely make the differences between /d/ tokens in CE and CF significant. Still, this pattern is different from the pattern observed for monolingual speakers in this study. For monolingual speakers, language differences are significant even when tokens without clear bursts are excluded from analysis.

Only one bilingual speaker produced language-specific differences in relative intensity values in CE and CF.⁵ Of the remaining four speakers, two speakers produced tokens in both languages with relative intensity similar to or lower than relative intensity of monolingual CE talkers, whereas two others produced tokens in both languages with relative intensity values similar to the relative intensity of monolingual CF talkers.

Mean burst frequency

The average mean burst frequencies for bilingual CE and CF speakers are summarized in Figure 3B. Overall, unlike the pattern observed for monolingual speakers, there was no consistent difference in mean frequency between CE and CF tokens produced by bilingual speakers. In the ANOVA, Only the main effect of Voicing ($F(1, 46) = 20.7, p < 0.01$) was significant. In CE as well as CF, /d/ tokens had a lower mean burst frequency than /t/ tokens; however, differences were significant only in CE. Similar to the pattern observed for monolingual speakers, bilingual speakers also produced mean frequency differences between /d/ and /t/ tokens in CE. However, unlike monolinguals, bilingual speakers did

⁵ An ANOVA of tokens produced by BF2 confirmed that tokens in CE and CF were significantly different in relative intensity (F -value for Language ($1, 11$) = 44, $p < 0.001$).

not produce mean frequency differences between /d/ and /t/ in CF.

Three bilingual speakers produced language-specific differences in CE and CF.⁶ However, the direction of difference between CE and CF tokens was reversed for speaker BM2; he produced coronal stops in CE with lower mean frequencies than in CF. Consequently, mean frequency differences did not reliably distinguish between CE and CF tokens. The remaining two speakers produced tokens in both languages with mean frequency similar to or higher than the mean frequency of monolingual CE talkers.

SD of burst frequency

The average SD of burst frequency for bilingual CE and CF speakers are summarized in Figure 4B. Overall, similar to the pattern observed for monolingual speakers, bilingual speakers consistently produced coronal stops in CE with lower SD of burst frequency than in CF. Thus, compared to CF coronals, spectra of CE coronals were compact with energy spread over a smaller range of frequencies. In the ANOVA, only the main effect of Language ($F(1, 46) = 69.8, p < 0.01$) was significant. Bonferroni's post-hoc tests confirmed that Language differences were significant for /d/ and /t/.

Four out of five bilingual subjects produced language-specific differences in SD of burst frequency in CE and CF.⁷ Even the fifth subject produced /d/ tokens differing in SD of burst frequency in the two languages. Thus, SD differences provided reliable cues to distinguish between CE and CF coronal stops produced by individual bilingual speakers as well.

Skewness of burst frequency

The average skewness of burst frequency for bilingual CE and CF speakers are summarized in Figure 5B. Overall, similar to the pattern observed for monolingual speakers, there was no consistent pattern of language difference either for group data or individual subject data for Skewness in CE and CF. There were no significant results in the ANOVA. Skewness was neither useful for distinguishing between CE and CF tokens, nor between voiced and voiceless tokens.

⁶ An ANOVA of tokens produced by BM1 ($F\text{-value}(1, 5) = 56, p < 0.01$), BM2 ($F\text{-value Language}(1, 9) = 21, p < 0.01$) and BF2 ($F\text{-value Language}(1, 11) = 11.8, p < 0.01$) confirmed that tokens in CE and CF were significantly different in mean frequency.

⁷ To confirm the observation, an ANOVA with Language (CE and CF) and Voicing (/d/ and /t/) as within subject variables was conducted on tokens produced by individual speakers. Speakers BM2 ($F\text{-value for Language}(1, 9) = 20.4, p < 0.01$), BF1 ($F\text{-value for Language}(1, 10) = 26.5, p < 0.001$), BF2 ($F\text{-value for Language}(1, 11) = 15.5, p < 0.01$) and BF3 ($F\text{-value for Language}(1, 7) = 55.8, p < 0.001$) produced significantly different SD for tokens across CE and CF.

Kurtosis of burst frequency

The average kurtosis of burst frequency for bilingual CE and CF speakers are summarized in Figure 6B. Overall, similar to the pattern observed for monolingual speakers, bilingual speakers consistently produced coronal stops in CE with higher kurtosis than in CF. However, in the ANOVA, the main effect of Language ($F(1, 46) = 3.9, p = 0.05$) was only marginally significant.

Two bilingual speakers produced language-specific differences in kurtosis in CE and CF.⁸ The remaining three bilingual speakers produced values intermediate to both monolingual groups. Thus, although kurtosis may help to differentiate between CE and CF tokens produced by bilingual speakers, it is clearly not as robust as SD at the group or individual level.

Summary

Analyses of data from bilingual speakers indicate that like monolingual speakers, bilingual speakers also produce language-specific differences in CE and CF. Tokens produced by bilinguals in the two languages were different in VOT and SD of burst frequency. They were also different on kurtosis and relative intensity; however, unlike the tokens produced by monolinguals, these differences were not large enough in the tokens produced by the bilinguals to reach significance. Furthermore, within CE, /d/ tokens were different from /t/ tokens in VOT, relative intensity and mean burst frequency, whereas within CF, /d/ tokens were different from /t/ tokens only in VOT.

Bilingual versus monolingual speakers

In this section, analyses are presented to address how (if at all) bilinguals differ from monolinguals in their production of coronal stops in the two languages. Group data from bilingual subjects in each language are compared to the respective monolingual group using t-tests to ascertain if bilinguals behave in CE and CF like monolinguals. In Table 2, results of t-tests are reported separately for /d/ and /t/ because voicing differences were significant for both monolingual and bilingual groups.

From Table 2 it is evident that bilingual speakers are similar but not identical to the monolingual speakers. Bilingual and monolingual speakers are more similar in the production of /t/ than /d/. Importantly, on the two variables on which both monolinguals and bilinguals produce language-specific differences – VOT & SD of burst frequency – there is a significant difference between the two groups (see highlighted rows in Table 2). Further, it

⁸ An ANOVA of tokens produced by BM2 ($F\text{-value Language}(1, 9) = 7.9, p < 0.05$) and BF3 ($F\text{-value Language}(1, 7) = 15.4, p < 0.01$) confirmed that tokens in CE and CF were significantly different in kurtosis.

Table 2. *T*-test results comparing monolingual and bilingual talkers on VOT, relative intensity, SD and kurtosis of burst frequency (* $p < 0.01$, ** $p < 0.001$). Standard deviations are in parentheses

| Variable | Canadian English | | | Canadian French | | |
|-------------------------|------------------|------------|--------|-----------------|------------|--------|
| | Monolingual | Bilingual | t | Monolingual | Bilingual | t |
| <i>/d/</i> | | | | | | |
| VOT (ms) | 8.4 (22) | -78.9 (59) | 12.5** | -90 (35) | -116 (26) | 4.9** |
| Relative Intensity (dB) | 12.7 (3) | 14.5 (7) | n.s. | 20.1 (4) | 15.4 (5) | 5.5** |
| SD (Hz) | 1929 (406) | 1704 (378) | 3.5* | 2470 (453) | 2298 (507) | n.s. |
| Kurtosis | 1.2 (1.8) | 2.3 (3.8) | n.s. | 0.04 (1.5) | 0.5 (1.6) | n.s. |
| <i>/t/</i> | | | | | | |
| VOT (ms) | 71.7 (21) | 68.7 (21) | n.s. | 20.7 (4) | 18.9 (6) | n.s. |
| Relative Intensity (dB) | 11.4 (5) | 11.9 (8) | n.s. | 16.1 (3) | 16.3 (8) | n.s. |
| SD (Hz) | 1538 (439) | 1690 (413) | n.s. | 2623 (485) | 2218 (500) | 5.3** |
| Kurtosis | 3.0 (3) | 1.4 (2) | 4.1** | -0.8 (1.3) | 1.5 (3) | -4.4** |

is evident from the VOT data that the bilinguals do not produce coronal stops in CE and CF with VOT values intermediate to that of monolinguals. Thus, bilinguals are different from L2 learners. Specifically, the differences between bilinguals and monolinguals are not consistent with a reduction or neutralization of voicing and place distinctions in the speech produced by bilingual speakers.

Discussion

In the present study, acoustic analyses of /d/ and /t/-initial words produced by simultaneous bilingual adults and monolingual CE and CF speakers in sentence context are reported. Simultaneous bilingual speakers as well as monolinguals produced language-specific differences to signal voicing differences in CE and CF. However, simultaneous bilinguals relied on only a subset of cues used by monolingual speakers of the two languages. Besides VOT, monolingual CE speakers produced /d/ and /t/ tokens differing in relative intensity, mean and SD of burst frequency; monolingual CF speakers produced /d/ and /t/ tokens differing in relative intensity and mean burst frequency. In contrast, simultaneous bilingual speakers produced /d/ and /t/ tokens in CE differing in VOT, relative intensity and mean frequency; they produced /d/ and /t/ tokens in CF differing only in VOT. Although voicing differences can be signaled by other acoustic cues (onset of fundamental frequency, onset of first formant frequency and length of following vowel) not measured in the present study, the differential use of burst amplitude and spectral cues to signal voicing highlights differences in phonetic implementation between simultaneous bilingual and monolingual speakers.

Acoustic analysis of place differences as measured by burst intensity and spectral cues also provides evidence for language-specific differences across the two languages.

Monolingual CE and CF speakers produced coronal stops in the two languages differing in relative intensity, mean, SD and kurtosis of burst frequency; simultaneous bilingual speakers produced coronal stops in the two languages differing in SD and also in relative intensity and kurtosis, although the latter differences were less robust. Again, similar to the pattern observed for voicing, bilingual speakers used only a subset of cues used by monolingual speakers to signal place differences.

In the following sections, first, findings from monolingual speakers are discussed in the context of language-specific differences between CE and CF. Then, the results from simultaneous bilingual speakers are integrated with previous research investigating speech production by children acquiring two languages from birth to address developmental issues relating to dual language exposure from birth. In the following section, findings from simultaneous bilingual speakers are compared with the data from early L2 learners to discuss effects of differences in age of exposure. Finally, data from simultaneous bilingual speakers are compared with data from monolingual adults to discuss organization of phonological systems in simultaneous bilingual adults.

Monolingual data

VOT results for coronal stop-initial words produced in sentences by monolingual CE, but not monolingual CF speakers, replicate analyses of tokens produced in isolation by Caramazza et al. (1973) and Sundara (2005). Consistent with previous reports, monolingual CE speakers produced non-overlapping distributions of VOT for /d/ and /t/. However, unlike previous reports, in a more naturalistic context – for tokens produced in sentences – there was no overlap in VOT for /d/ and /t/ produced by monolingual CF speakers.

There are two, mutually non-exclusive, explanations to account for the differences in the VOT distributions reported in this paper and the data reported by Caramazza et al. (1973) and Sundara (2005). First, target word-initial stops in the sentence frame were preceded by a voiced segment, namely the vowel /i/ in the word *dis*, that is more conducive to voicing during stop closure. Thus, the phonetic context in the sentence context may account for a greater number of CF /d/ tokens being produced with lead rather than short-lag VOT.⁹ Second, overall shorter syllable durations for the tokens produced in isolation as compared to the tokens produced in the sentences may also account for the overlap in VOT observed in the former case. This was certainly the case when syllable durations from this paper were compared to that from tokens produced by the same speakers in isolation (analyses for the latter are reported in Sundara, 2005). Shorter syllable durations for repeated iterations of the same word when compared to the first instance have been previously reported (Fowler and Housum, 1987). Fowler and Housum (1987) argue that the longer syllable duration of the first utterance serves as a referent for subsequent iterations. Recall that target words were first recorded in sentence context, and hence longer than tokens produced in isolation. Also, when reading items on a list, tokens in utterance final position may be produced with a faster rate of speech, which, in turn, results in shorter syllable durations overall.

The non-overlapping distribution of VOT for /d/ and /t/ tokens produced in CF sentences calls into question previous reports of dialectal differences in VOT implementation in CF and French from France (European French or EF). In a comparison of VOT patterns produced by CF and EF speakers, Caramazza and Yeni-Komshian (1974) report that unlike CF, EF speakers produce non-overlapping VOT values associated with voiced and voiceless stops. Typically, voiced stops in EF are produced with lead VOT whereas voiceless stops are produced with a short-lag VOT. Caramazza and Yeni-Komshian (1974) attribute differences in the two dialects of French to the development of CF with extensive contact with CE. VOT patterns for /d/- and /t/-initial words produced in sentences produced by CF speakers in this study are identical to the patterns described by Caramazza and Yeni-Komshian (1974) for EF speakers. In view of the results from this study, VOT distributions produced by EF and CF speakers need to be compared controlling for overall differences in syllable duration and phonetic context to make strong claims about phonetic change relating to language contact.

⁹ We would like to thank Susan Guion for pointing out the difference in phonetic context for tokens produced in isolation and in the sentence context.

Bilingual data

Developmental consequences of dual language exposure

Recall that developmental data from simultaneous bilingual children indicate that these children produce language-specific differences in the production of /t/ but not /d/ (Watson, 1990; Khattab, 2000; Johnson and Wilson, 2002), primarily because bilingual children do not produce /d/ tokens with lead VOT even by age 10. VOT data reported in this study confirm that adult bilinguals exposed to two languages from birth indeed produce language-specific differences for /t/ as well as /d/ in their two languages. Furthermore, by adulthood, simultaneous bilinguals have mastered production of lead VOT, suggesting that differences in VOT production for voiced stops observed between bilingual and monolingual children are likely to reflect a difference in rate of acquisition.

Effects of age of exposure

To evaluate the effects of age of exposure on phonetic implementation when two groups of speakers are exposed to dual language input, phonetic implementation by simultaneous bilingual adults is compared in this section to the patterns reported in the literature for L2 learners. A comparison of data from acoustic analyses of speech produced by adult simultaneous bilingual speakers tested in this study and L2 learners reveals similarities as well as differences. Language-specific differences in VOT of /t/ tokens reported here for simultaneous bilingual adults have also been reported by Caramazza et al. (1973) and others (Williams, 1979; Flege and Eefting, 1987a; Flege, 1991; Grosjean and Miller, 1994) for early L2 learners. However, although early L2 learners produce language-specific differences in VOT in their two languages having learned to produce long-lag VOT in their L2, they still remain significantly different from monolingual speakers. Recall that Caramazza et al. report that early L2 learners produced /t/ in French like monolingual CF speakers, but they produced /t/ in English with VOT intermediate to monolingual CE and CF speakers. This was clearly not the case for simultaneous bilingual adults recorded in this study. In CE as well as in CF, VOT for /t/ tokens produced by bilinguals (Mean = 68.7, SD = 21; 18.9, 6) and monolinguals (71.7, 21; 20.7, 4) was virtually identical. Therefore, early exposure seems crucial for the development of native-like VOT values for /t/ tokens in CE and CF.

Although simultaneous bilingual adults and early L2 learners may differ in language use patterns as well as the nature of their early input, recent research suggests that native-like VOT values for voiceless stops can be established when there is early input even in the absence of regular use of the two languages (Au, Knightly,

Jun and Oh, 2002). In their study, Au and colleagues compared English-speaking adults who had been exposed to Spanish for about nine hours a week for at least three years between birth and age 6 years with adult L2 learners of Spanish without this early exposure. Au et al. (2002) report that compared to adult L2 learners, childhood overhearers had more native-like accents when speaking Spanish. They also produced /t/ tokens with short-lag VOT values comparable to productions by native Spanish speakers. Thus, early effects of speech input on VOT of voiceless stops may persist up to adulthood, even in cases when exposure has been discontinued.

Data on production of lead VOT by L2 learners exposed to a second language early in life are scarce for a variety of reasons. First, most studies investigate L2 learners of English, and voiced tokens in English are typically produced with short-lag VOT. Second, even within monolingual English speakers, there is considerable variability in the production of voiced tokens. Several researchers have reported that in American English, monolingual speakers also produce voiced tokens with lead VOT (Flege and Eefting, 1987b; Mack, 1989), possibly due to idiolectal and dialectal variations between subject populations. Consequently, in English, voiced tokens with either lead or short-lag VOT are considered native-like. Third, although native English listeners have been demonstrated to be sensitive to small deviations in long-lag VOT in voiceless stops produced by non-native speakers (Flege and Hammond, 1982; Flege, 1984), there is no evidence to indicate that native English listeners are able to detect differences in lead VOT in voiced stops.

Although Caramazza et al. (1973) do not report averages for /d/ for CF speakers learning CE, given that their L2 learners were native speakers of CF and /d/ tokens in CF are produced with lead VOT, production of lead VOT is unlikely to be problematic. This is borne out in the frequency distribution of VOT values presented in Caramazza et al; L2 learners produced voiced tokens in CE and CF with lead VOT. Hazan and Boulakia (1993) provide data from English speaking adults who had started learning French before 5 years of age. They report that early L2 learners of French produced voiced tokens with lead VOT in both languages, but differentiated between lead VOT for English and French /b/.

Simultaneous bilingual speakers in this study also produced voiced tokens with lead VOT in both languages. Also, like the early L2 learners (Hazan and Boulakia, 1993), simultaneous bilinguals produced /d/ in CE and CF with significantly different lead VOT values. Given that early L2 learners as well as simultaneous bilingual adults are able to produce voiced tokens with lead VOT, it is unlikely that production of lead VOT is determined by early experience alone. Thus, the effects of early exposure are evident in the production of some, but not all segments.

Phonological organization in bilinguals

The similarities and differences in production of VOT by L2 learners and monolinguals are particularly pertinent when discussing underlying phonological systems of L2 learners. Inferences about the phonological system of L2 learners in production studies require two pieces of data. First, language-specific differences in the production of the phonetic segments in the two languages by L2 learners must be established to provide necessary evidence for two systems of representation, one for each language. Then, phonetic implementation by L2 learners must be compared to implementation by monolingual or native controls to determine whether the two systems are independent. Specifically, in the psycholinguistic and L2 research tradition, differences between L2 learners and monolinguals that cannot be attributed to input alone are considered a consequence of cross-language influences or interactions.

Recall that the previously available VOT data from children exposed to two languages from birth are consistent with two language systems for /t/ but not /d/ (Watson, 1990; Khattab, 2000; Johnson and Wilson, 2002). In this study, to optimize the likelihood of obtaining language-specific differences, simultaneous bilingual subjects were recorded in two sessions, each conducted only in CE or in CF by different testers. The data from simultaneous bilingual adults as well indicates language-specific production of coronal stops in CE and CF providing the necessary evidence for two systems, one for each language. We demonstrate this for voicing as well as place differences in coronal stops.

Although simultaneous bilinguals and monolinguals produce language-specific differences in coronal stops in CE and CF, they differ in two ways. First, simultaneous bilinguals use only a subset of acoustic cues used by monolinguals to differentiate coronal stops in CE and CF. This is evident from the comparison of group data as well as in the performance of individual simultaneous bilingual subjects. Second, even when simultaneous bilinguals and monolinguals produce language-specific differences in CE and CF on the same variables, they are significantly different. This is evident from t-test comparisons of VOT, relative intensity, SD and Kurtosis of burst frequency presented in Table 2. Unless input received by bilinguals can be characterized as qualitatively different from that received by monolinguals, the differences in phonetic implementation between bilingual and monolingual speakers cannot be reconciled as relating to language-experience alone.

The differences between phonetic implementation by simultaneous bilinguals and monolinguals provide evidence for cross-language influences or interactions between the two languages of the simultaneous bilinguals. Although interactions between the two languages have been explored in detail in the case of L2 learners

(Grosjean, 1989; Best, 1995; Flege et al., 2003), the idea of interactions between the languages of simultaneous bilinguals is fairly new. More importantly, this study provides the first evidence that interactions between the languages of simultaneous bilinguals are not developmentally transcribed and hence, temporary; these interactions continue to be evident even in the productions of adult simultaneous bilinguals.

The differences between bilinguals and monolinguals can also be thought to model the effects engendered by language contact within an individual producer/perceiver that may serve as precursors for language change. Thus, it is likely that cross-language interactions are governed by universal constraints on organization of speech perception/production behavior and human learning mechanisms. Specifically, phonetic implementation by bilinguals may reflect an economy in the use of phonetic and articulatory routines, as suggested by Watson (1990, 1996), or a strategy to make perceptual distinctions more salient, or some combination of the two. Several researchers have previously suggested that perceptual distinctiveness and articulatory ease play a crucial role in modeling language change and in explaining language acquisition (Liljencrants and Lindblom, 1972; Ohala, 1978, 1990; Lindblom, 1983, 1990; Ohala and Ohala, 1993; Boersma, 1998).

The results from this study support the idea that coronal stops produced by bilingual speakers may be perceptually more distinct than coronal stops produced by monolingual speakers. Consider the VOT differences between /d/ and /t/ tokens produced by monolingual and bilingual speakers. The difference between VOT for /d/ and /t/ tokens produced by bilinguals in CE (147.6 ms) as well as CF (134.9 ms) was greater than the difference between VOT for /d/ and /t/ produced by monolingual CE (63.3 ms) or CF (110.7 ms) speakers. Because greater differences in VOT between voiced and voiceless tokens are more likely to be detected than smaller ones, it can be argued that voicing differences in coronal stops produced by bilinguals are more perceptually salient than voicing differences produced by monolinguals.

Cues for place differences, however, cannot be compared as easily as VOT differences because these comparisons need to be made across speakers for the monolingual group whereas for the simultaneous bilingual group, the comparison is within-speaker. Researchers have previously suggested that as a direct consequence of having one or other (but not both) subgroup of coronal stops in the phonetic inventory, English speakers are more variable while producing English /d/ (Jongman et al., 1985; Stevens, Keyser and Kawasaki, 1985). Dental allophones of /d/ and /t/ occur in English, specifically preceding inter-dental consonants; some researchers have claimed that coronal consonants are dentalized in several dialects of English (Francis, 1958) or even that some

English speakers do not distinguish between dental and alveolar stops, often interchanging them (Dixon, 1980). It may be argued that because simultaneous bilinguals are exposed to two sub-groups of coronal stops, they produce more distinct distributions of coronal stops when compared to monolinguals. A comparison of production of coronal stops by simultaneous bilingual speakers and speakers of Australian languages where coronal place distinctions contrast meaning are needed to test this hypothesis.

The similarities in phonetic production by different kinds of language learners, on the other hand, reflect robust effects of specific language-experience. For production of segments, speakers often rely on multiple cues, some of which are necessary, some sufficient and still others redundant. The cues used consistently by different kinds of native speakers, bilinguals and monolinguals, can help identify the core cues that constitute native-like behavior and also isolate cues that are less relevant in signaling phonetic contrasts. Therefore, the similarities between bilinguals and monolinguals inform us about acquisition by narrowing the aspects of segmental perception and production that emerge from native language exposure alone.

Recall that native CE and CF listeners did not rate spontaneous speech samples produced by monolingual and bilingual speakers differently indicating that native listeners did not perceive differences in phonetic implementation by monolingual and simultaneous bilingual speakers. Arguably, because simultaneous bilinguals produce coronal stops in the two languages differing on fewer acoustic cues than monolinguals but with no perceptible consequences, bilinguals' productions of coronal stops in the two languages can be considered more streamlined than the productions of monolinguals. Thus, differences in VOT and SD may be necessary and sufficient to characterize native-like productions of /d/ and /t/ in CE and CF.

In the absence of articulatory data, it remains unclear what the articulatory gestures underlying coronal stop production in CE and CF are. Besides place of articulation differences which are manifested as differences in relative burst intensity, researchers have suggested that sub-groups of coronal stops differ in the length of constriction (Chomsky and Halle, 1968) or the active articulator used to make the constriction (Stevens et al., 1985). Future research is needed to determine how differences in place, length of constriction, and active articulator (tongue tip or tongue blade) relate to mean, SD, skewness and kurtosis of burst frequency (for a more detailed discussion of hypothesized articulatory correlates of spectral moments see Sundara (2005)).

Although the subjects in this study were exposed to both English and French at home from birth, no claim is made as to whether their input was balanced across

their two languages. As De Houwer and Bornstein (2003) report, even in “one-parent one-language” families, the amount of input received by infants in each of the two languages is seldom 50–50. As indicated, the five subjects were chosen such that they were using both languages regularly in the home and work context and were heard as being native-like by monolingual listeners. Clearly a greater number of simultaneous bilinguals do not fit these criteria than do. Thus, no claim is made about how bilinguals typically behave. Such a claim would require testing of much larger sample sizes of simultaneous bilingual individuals than has been done here.

What the study does demonstrate unequivocally is that even when bilinguals sound like monolinguals, they differ from each other in implementation of phonetic segments, quite possibly to highlight perceptual differences or for articulatory ease. What remains to be confirmed is if, indeed, accommodations made by bilinguals are more perceptually distinct or easier to produce articulatorily. In any case, the documentation of effects of dual language exposure from birth provides an essential baseline for investigations of development of simultaneous bilinguals and effects of age of L2 exposure on production of place and voicing differences.

References

- Au, T. K., Knightly, L. M., Jun, S. & Oh, J. S. (2002). Overhearing a language in childhood. *Psychological Science*, 13, 238–243.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, pp. 167–200. Timonium, MD: York Press.
- Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.
- Boersma, P. & Weenink, D. (2005). Praat: Doing phonetics by computer (Version 4.3.19) [Computer program]. Retrieved July 20, 2005, from <<http://www.praat.org/>>.
- Caramazza, A. & Yeni-Komshian, G. H. (1974). Voice onset time in two French dialects. *Journal of Phonetics*, 2, 239–245.
- Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B. & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French–English bilinguals. *Journal of Acoustical Society of America*, 54, 421–428.
- Celce-Murcia, M. (1978). The simultaneous acquisition of English and French in a two-year-old child. In E. Hatch (ed.), *Second language acquisition: A book of readings*, pp. 38–53. Rowley, MA: Newbury House.
- Chomsky, N. & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.
- Cole, J., Choi, H., Kim, H. & Hasegawa-Johnson, M. (2003). The effect of accent on the acoustic cues to stop voicing in Radio News Speech. In M. J. Solé, D. Recasens & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, pp. 2665–2668.
- De Houwer, A. & Bornstein, M. H. (2003). Balancing on the tightrope: Language use patterns in bilingual families with young children. Presented at 4th International Symposium on Bilingualism, Tempe, AZ.
- Dixon, R. M. W. (1980). *The languages of Australia*. Cambridge: Cambridge University Press.
- Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America*, 76, 692–707.
- 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 Acoustical Society of America*, 89 (1), 395–411.
- Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, pp. 233–273. Timonium, MD: York Press.
- Flege, J. E. & Eefting, W. (1987a). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, 15, 67–83.
- Flege, J. E. & Eefting, W. (1987b). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communications*, 6, 185–202.
- Flege, J. E. & Hammond, R. (1982). Mimicry of non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition*, 5, 1–18.
- Flege, J. E., Munro, M. J. & MacKay, I. R. A. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication*, 16, 1–26.
- Flege, J. E., Schirru, C. & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, 40, 467–491.
- Flege, J. E., Takagi, N. & Mann, V. (1995). Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech*, 38, 25–55.
- Forrest, K., Weismer, G., Milenkovic, P. & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America*, 84 (1), 115–123.
- Fowler, C. A. & Housum, J. (1987). Talkers’ signaling of “new” and “old” words in speech and listeners’ perception and use of the distinction. *Journal of Memory and Language*, 26, 489–504.
- Francis, W. N. (1958). *The structure of American English*. New York: Ronald Press Co.
- Genesee, F., Boivin, I. & Nicoladis, E. (1996). Talking with strangers: A study of bilingual children’s communicative competence. *Applied Psycholinguistics*, 17, 442–447.
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36, 3–15.
- Grosjean, F. (1997). Processing mixed language: Issues, findings and models. In A. M. B. de Groot & J. F. Kroll (eds.), *Tutorials in Bilingualism: Psycholinguistic perspectives*, pp. 225–254. Mahwah, NJ: Lawrence Erlbaum Associates.
- Grosjean, F. & Miller, J. L. (1994). Going in and out of languages: An example of bilingual flexibility. *Psychological Science*, 5, 201–206.

- Guion, S. G. (2003). The vowel systems of Quichua–Spanish bilinguals: Age of acquisition effects on the mutual influence of the first and second languages. *Phonetica*, 60, 98–128.
- Hazan, V. L. & Boulakia, G. (1993). Perception and production of a voicing contrast by French–English bilinguals. *Language and Speech*, 1, 17–38.
- Homma, Y. (1980). Voice onset time in Japanese stops. *Onsei Gakka Kaiho*, 163, 7–9. [New journal title: *Journal of the Phonetic Society of Japan*.]
- Johnson, C. E. & Lancaster, P. (1998). The development of more than one phonology: A case study of a Norwegian–English bilingual child. *International Journal of Bilingualism*, 2, 265–300.
- Johnson, C. E. & Wilson, I. L. (2002). Phonetic evidence for early language differentiation: Research issues and some preliminary data. *The International Journal of Bilingualism*, 6, 271–289.
- Jongman, A., Blumstein, S. E. & Lahiri, A. (1985). Acoustic properties for dental and alveolar stop consonants: A cross-language study. *Journal of Phonetics*, 13, 235–251.
- Kent, R. D. (1992). The biology of phonological development. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (eds.), *Phonological development: Models, research, implications*, pp. 65–90. Timonium, MD: York Press.
- Khattab, G. (2000). VOT production in English and Arabic bilingual and monolingual children. *Leeds Working Papers in Linguistics*, 8, 95–122.
- Lanza, E. (1992). Can bilingual 2-year-olds code-switch? *Journal of Child Language*, 19, 633–658.
- Leopold, W. (1949/1970). *Speech development of a bilingual child: A linguist's record* (vols. 1–4). New York: AMS Press.
- Lieberman, P. & Blumstein, S. E. (1988). *Speech physiology, speech perception, and acoustic phonetics*. New York: Cambridge University Press.
- Liljencrants, J. & Lindblom, B. (1972). Numerical simulations of vowel systems: The role of perceptual contrast. *Language*, 48, 839–862.
- Lindblom, B. (1983). The economy of speech gestures. In P. F. MacNeilage (ed.), *The production of speech*, pp. 217–245. New York: Springer-Verlag.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H & H theory. In W. J. Hardcastle & A. Marchal (eds.), *Speech production and speech modeling*, pp. 403–439. Dordrecht: Kluwer.
- Lisker, L. (1978). Rapid vs. rabid: A catalogue of acoustic features that may cue the distinction. *Haskins Laboratory Status Report*, 54, 127–132.
- Lisker, L. & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384–422.
- Mack, M. (1989). Consonant and vowel perception and production: Early English–French bilinguals and English monolinguals. *Perception and Psychophysics*, 46, 187–200.
- Macken, M. & Barton, D. (1980). The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. *Journal of Child Language*, 7, 433–458.
- Mayer, M. (1969). *Frog, where are you?* New York: Puffin Pied Piper.
- Ohala, J. J. (1978). Production of tone. In V. Fromkin (ed.), *Tone: A linguistic survey*, pp. 5–39. New York: Academic Press.
- Ohala, J. J. (1990). The phonetics and phonology of aspects of assimilation. In J. Kingston & M. Beckman (eds.), *Papers in laboratory phonology*, vol. 1: *Between grammar and the physics of speech*, pp. 258–275. Cambridge: Cambridge University Press.
- Ohala, J. J. & Ohala, M. (1993). The phonetics of nasal phonology: Theorems and data. In M. K. Huffman & R. A. Krakow (eds.), *Nasals, nasalization and the velum*, pp. 225–249. San Diego: Academic Press.
- Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of Acoustical Society of America*, 75, 224–230.
- Picard, M. (1987). *An introduction to the comparative phonetics of English and French in North America*. Amsterdam & Philadelphia: John Benjamin.
- Picard, M. (2001). *Phonetics and phonology for ESL and TESL teachers: Comparing Canadian English and French*. Montreal: Concordia University.
- Schnitzer, M. L. & Krasinski, E. (1994). The development of segmental phonological production in a bilingual child. *Journal of Child Language*, 21, 585–622.
- Schnitzer, M. L. & Krasinski, E. (1996). The development of segmental phonological production in a bilingual child: A contrasting second case. *Journal of Child Language*, 23, 547–571.
- Stevens, K. N., Keyser, S. J. & Kawasaki, H. (1985). Toward a phonetic and phonological theory of redundant features. In J. S. Perkell & D. H. Klatt (eds.), *Invariance and variability of speech processes*, pp. 426–463. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Stoel-Gammon, C., Williams, K. & Buder, E. (1994). Cross-language differences in phonological acquisition: Swedish and American /t/. *Phonetica*, 51, 146–158.
- Summerfield, Q. & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of Acoustical Society of America*, 62, 435–448.
- Sundara, M. (2005). Acoustic-phonetics of coronal stops: A cross-language study of Canadian–English & Canadian–French. *Journal of the Acoustical Society of America*, 118 (2), 1026–1037.
- Treffers-Daller, J. (1997). Variability in code switching styles: Turkish–German code switching patterns. In R. Jakobson (ed.), *Code-switching worldwide*, pp. 177–197. Berlin: Mouton de Gruyter.
- Vogel, I. (1975). One system or two: An analysis of a two-year-old Romanian–English bilingual's phonology. *Papers and Reports on Child Language Development*, 9, 43–62.
- Watson, I. (1990). Acquiring the voicing contrast in French: A comparative study of monolingual and bilingual children. In J. N. Green & W. Ayers-Bennett (eds.), *Variation and*

- change in French: Essays presented to Rebecca Posner on the occasion of her sixtieth birthday*, pp. 37–60. London: Routledge.
- Watson, I. (1996). Phonological processing in two languages. In E. Bialystok (ed.), *Language processing in bilingual children*, pp. 25–48. Cambridge: Cambridge University Press.
- Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception and Psychophysics*, 26, 95–105.
- Yeni-Komshian, G. H., Caramazza, A. & Preston, M. S. (1977). A study of voicing in Lebanese Arabic. *Journal of Phonetics*, 5, 35–48.
- Zlatin, M. & Koenigsknecht, R. (1976). Development of the voicing contrast. *Journal of Speech and Hearing Research*, 19, 92–111.

Received May 12, 2004

Revision received April 18, 2005 and July 6, 2005

Accepted July 18, 2005