



Cross language phonetic influences on the speech of French–English bilinguals

Carol A. Fowler^{a,b,*}, Valery Sramko^c, David J. Ostry^{a,c},
Sarah A. Rowland^{a,b}, Pierre Hallé^{a,d}

^a*Haskins Laboratories, 300 George St., Suite 900, New Haven, CT 06511, USA*

^b*University of Connecticut, 406 Babbodge Road, Unit 1020 Storrs, CT 06269, USA*

^c*McGill University, 1205 Dr. Penfield Ave., Montreal, QC H3A 1B1, Canada*

^d*Laboratoire de Phonétique et Phonologie, CNRS-Paris 3, France*

Received 21 August 2007; received in revised form 29 January 2008; accepted 14 April 2008

Abstract

We examined the voice onset times (VOTs) of monolingual and bilingual speakers of English and French to address the question whether cross language phonetic influences occur particularly in simultaneous bilinguals (i.e., speakers who learned both languages from birth). Speakers produced sentences in which there were target words with initial /p/, /t/ or /k/. In French, natively bilingual speakers produced VOTs that were significantly longer than those of monolingual French speakers. French VOTs were even longer in bilingual speakers who learned English before learning French. The outcome was analogous in English speech. Natively bilingual speakers produced shorter English VOTs than monolingual speakers. English VOTs were even shorter in the speech of bilinguals who learned French before English. Bilingual speakers had significantly longer VOTs in their English speech than in their French. Accordingly, the cross language effects do not occur because natively bilingual speakers adopt voiceless stop categories intermediate between those of native English and French speakers that serve both languages. Monolingual speakers of French or English in Montreal had VOTs nearly identical respectively to those of monolingual Parisian French speakers and those of monolingual Connecticut English speakers. These results suggest that mere exposure to a second language does not underlie the cross language phonetic effect; however, these findings must be resolved with others that appear to show an effect of overhearing.

© 2008 Elsevier Ltd. All rights reserved.

0. Introduction

Talkers well past the end of any critical period for language acquisition show considerable malleability in their production of phonological segments. In the short term, for example, they may show convergences or accommodations to their interlocutors in dialect (Giles, 1973), speaking rate (Street, 1983), vocal intensity (Natale, 1975), and rate and duration of pausing (Jaffe & Feldstein, 1970) among other vocal adjustments (see Giles, Coupland, & Coupland, 1991, for a review to that date). These accommodations can be accompanied

*Corresponding author at: Haskins Laboratories, 300 George St., Suite 900, New Haven, CT 06511, USA. Tel.: +1 203 865 6163x220/+1 203 432 0890; fax: +1 203 865 8963.

E-mail address: carol.fowler@haskins.yale.edu (C.A. Fowler).

by nonverbal accommodations to interlocutors, such as posture mirroring (LaFrance, 1982), interactional synchrony (Condon, 1976), and entrainment of postural activity (Shockley, Santana, & Fowler, 2003). These short-term changes in vocal and nonvocal behavior may reflect interlocutors' efforts to coordinate their activities in the service of conjoint goals (e.g., Clark, 1996), a speculation supported by findings of divergence in vocal behavior under conditions in which interlocutors may experience hostility toward one another (e.g., Bourhis & Giles, 1977; Labov, 1963).

Somewhat (Sancier & Fowler, 1997) and considerably (Flege, 1987) longer term reflections of learning well past any critical period also occur. In an exploration of the speech of a single bilingual speaker of Brazilian Portuguese (her first language or L1) and English (L2, learned beginning at age 15 years), Sancier and Fowler found changes in the voice onset times (VOTs) of the speaker's voiceless stops in both languages as a function of the ambient language. Portuguese has uniformly unaspirated (short-lag) voiceless stops; in stressed, syllable-initial position, English has aspirated (long-lag) voiceless stops. When the bilingual speaker's speech was recorded after she had spent several months in Connecticut speaking English almost exclusively, her VOTs in both Portuguese and English were significantly longer than in recordings collected immediately after a 2-month stay in Brazil. Of interest in this study was the finding that parallel changes took place in the speaker's two languages even though she was a "serial bilingual"—that is, she spoke only English in the US and only Portuguese in Brazil. This suggests a psychological link between similar categories (compare, e.g., Flege, 1995) in the two languages—that is, categories in the two languages that bilingual speakers consider to be variants of the same phonological segment.

Flege (1987) reported evidence that phonological learning among adults occurs in the longer term still. He studied the VOTs of native speakers of French or English who had spent, on average, about 12 years in Chicago and Paris, respectively. Four of the seven French speakers in Chicago were married to native speakers of English; for all of them, English was their principal language. All of the native English speakers in Paris were married to native French speakers, and French was their principal language. It is not surprising that, among the native French speakers in Chicago, VOTs of English /t/ were shorter than those of monolingual English speakers (49 ms vs 77 ms, respectively). It is of greater interest that the VOTs of their French /t/ were longer than those of monolingual French speakers (51 ms vs 33 ms). Findings from English speakers in Paris were complementary. Their French VOTs were longer than those of monolingual French speakers (43 ms vs 33 ms). Their English VOTs were shorter than those of monolingual English speakers (49 ms vs 77 ms). It is less clear in these data than in those of Sancier and Fowler that the speakers had distinct categories in their two languages. Another difference in the data is the probable magnitude of the shifts in VOT. If we estimate the French speakers' "before-immersion" VOTs from those of the monolingual French speakers, their French VOTs shifted by about 18 ms on average. English speakers shifted by 28 ms. This is considerably more than the 5–6 ms shifts observed by Sancier and Fowler, and it probably reflects the much longer immersion times (12 years vs 2–4 months). Thus, phonological learning can occur, probably over a period of years, and the learning can affect production of the well-established L1 as well as the less entrenched L2.

In the studies by Flege (1987) and Sancier and Fowler (1997), parallel changes occurred in L1 and L2. For example, when Sancier and Fowler's participant, in an English-speaking environment, produced aspirated voiceless English stops with longer VOTs than when she was in a Portuguese environment, her unaspirated voiceless stops of Portuguese also lengthened. In order for parallel changes to be identifiable as such, the two languages must have categories that are, somehow, identified with one another. Flege (1995) refers to this as "equivalence classification." That is, although the (short-lag) voiceless stops of Portuguese and French are not identical to the (long lag) voiceless stops of English, speakers appear to treat them as variants of one another that are, in some way, linked cognitively.

Changes in phone production in L1 or L2 are not always assimilatory in nature. For example, Flege and Eefting (1987) examined the unaspirated, short-lag VOTs of Dutch voiceless stops and the aspirated, long lag VOTs of English among speakers for whom Dutch was the first language and English the second. Speakers varied in English proficiency, and Flege and Eefting found *shorter* VOTs in Dutch voiceless stops of speakers who were more, as compared to less, proficient speakers of English. That is, as their proficiency in English grew, they produced unaspirated Dutch stops that were dissimilated, and thus more different from the aspirated English stops, than productions of native Dutch speakers who were less proficient in English.

Flege and Eefting (1987) interpreted both findings, those of assimilation of similar segments in L1 and L2, and those of dissimilation, in terms of Flege's Speech Learning Model (SLM, e.g., Flege, 1995). In the model, L2 learners may or may not notice phonetic differences among similar phones in two languages. Formation of a distinct category for an L2 phone will be blocked by equivalence classification if phones are sufficiently similar. This does not mean that the phones in the two languages will be produced identically, however. The language user may notice, for example, a difference in VOT distributions. In any case, if a new category is not formed, production of phones in one or both languages will assimilate. If, in contrast, a new category for an L2 phone *is* established—that is, if equivalence classification does not occur—an L1 category (such as the voiceless stop categories of Dutch in the study by Flege and Eefting (1987) may be deflected away from that of monolingual speakers of L1 to maintain a contrast with the similar L2 phone category.

The preceding studies and many others show the malleability of production (and perception) of phonetic categories among language users who are well beyond any critical period for language acquisition. Malleability may be observed because the L2 is in the process of being learned or (for example in the studies by Sancier and Fowler (1997) and Flege (1987)), because the speaker's language environment changes. The malleabilities that are observed suggest that consequences of learning that support L1 and L2 language use are not independent.

What of mature speakers who are simultaneous bilinguals in a stable language environment? That is, what of speakers who learn two languages from birth and who are not subject to the sources of malleability just described? Do their language systems achieve a measure of independence so that their speech in both languages is like that of monolingual speakers? This is the first question that our research addresses.

Relatedly, Mack (1989) tested the production and perception in English of /d-/t/ and /i-/ɪ/. Bilingual speaker–listeners learned both French and English at an early age, but considered themselves more proficient in English. Their perceptions and productions of stops and vowels were compared to those of monolingual speakers. Mack found very limited evidence of differences between the groups. In perception, bilingual listeners had less steep identification functions for /d-/t/ (which differed in VOT) than monolingual listeners and an identification boundary along the /i-/ɪ/ continuum closer to the /i/ end of the continuum. (This is consistent with an effect of their experience with French, which has a more peripheral /i/ than English.) However, otherwise identification functions were the same for the two groups, and there were no group differences in discrimination. In production, there were almost no differences at all. The only notable difference occurred in /i/ productions where bilingual speakers produced a decrease of 50 Hz or more in F2 from vowel mid- to endpoint with greater frequency than did monolingual speakers. These findings suggest that early bilinguals may have memory systems for their two languages that are substantially, although not entirely, independent of one another.

Guion (2003) found largely compatible findings among native speakers of highland Ecuadorian Quichua and Spanish. Her speakers learned Spanish at various ages; however, of particular interest here are the native speakers of both languages. The dialect of Quichuan spoken by her speakers has three vowels transcribed phonetically as [ɪ], [ʊ] and [a]. Spanish has five vowels [i], [e], [o], [a], and [u]. Guion found that four of five simultaneous bilingual speakers distinguished Quichuan [ɪ] from both Spanish [i] and [e] in their productions. Only one of those speakers distinguished Quichuan [ʊ] from both Spanish [o] and [u]; the others distinguished the Quichuan vowel from one of the two Spanish vowels. In a further statistical comparison, she found that simultaneous bilinguals speaking Spanish produced vowels that were not distinct from those of monolingual Spanish speakers. Accordingly, as Mack (1989) had found for early bilinguals, the simultaneous bilinguals of this study had phonological systems that were largely, but not entirely, independent in the two languages.

Sundara, Polka and Baum (2006) identify Guion's (2003) study as the only one to date that has provided acoustical assessments of the speech of simultaneous bilinguals (Mack's speakers being "early" bilinguals). They provided another assessment, closely related to our study. They provided acoustical comparisons of the speech of six monolingual speakers of Canadian English, of six monolingual Canadian French speakers and of five simultaneous Canadian English and French speakers. As in our study, they examined the voiceless stops in the two languages. They provided a number of measures, including the measure on which we focus, VOT.

They found a number of measures on which monolingual speakers of English and French differed in their productions of /d/ and /t/. Whereas the French speakers prevoiced all of their productions of /d/, 95% of monolingual English speakers' /d/s were produced with a short lag VOT. Whereas 100% of productions of /t/

by monolingual French speakers had short lag VOTs, 100% of English tokens had long lag VOTs (greater than 30 ms). English speakers had lower relative burst intensities (that is, maximum vowel intensity minus burst intensity) than French speakers, higher mean burst frequencies. Smaller standard deviations of burst frequencies, no consistent language differences in skewness, and higher kurtosis in English.

Comparing the simultaneous bilinguals speaking English and French with respective monolingual speakers, they found a few differences. First, bilinguals produced English /d/ with prevoicing on 74% of tokens, considerably more frequently than monolingual English speakers. Their English /t/ productions and French /t/ and /d/ were like those of monolingual speakers in VOT. In respect to relative burst intensity, they showed the same language difference as the monolingual speakers, but for /t/ only; relative intensity was the same for French and English /d/. In contrast to monolingual speakers, bilingual speakers showed no difference in burst frequency across the languages. They showed nearly the same language differences or lack thereof in respect to standard deviation, skewness, and kurtosis of burst frequency.

In our research, we will extend this investigation to a larger sample of monolingual and bilingual speakers to investigate VOT production further. Our main interest, like that of Sundara, et al., is in simultaneous bilinguals; however, we include samples of English- and French-L1 bilinguals as well. A focus on VOT may appear an odd choice given the findings of Sundara et al. that simultaneous bilinguals did not differ from English and French monolinguals on this measure (in production of /d/ and /t/). In that study, both English speaking groups showed long lag VOTs on /t/ that were nonsignificantly different in duration, and both French speaking groups showed uniformly short lags in production of /t/ that did not differ significantly in duration.

We chose to look at voiceless VOTs, because of our findings in the study by Sancier and Fowler (1997) in which our speaker, a mid to late (age 15 years) learner of English, showed significant changes in her English and Portuguese voiceless VOTs due to changes in the language environment. These differences, though small (5–6 ms), were statistically significant and were audible in her Portuguese speech as judged by native listeners. Although there are many differences between that study and the study of Sundara et al. (e.g., a sequential vs simultaneous bilinguals, transient vs enduring differences in experienced voiceless stop aspiration due to language environment), none of them explain to our satisfaction why VOT was malleable in our study, but not in that of Sundara et al. Accordingly, we chose to take another look with a larger sample size.

We were interested in a second issue in addition to that of the independence of the language systems of simultaneous bilinguals as indexed by VOT. A few studies have asked whether exposure to a language that one does not speak or understand has an impact either on learning the language of exposure at a later date or on production of the native language.

As to the former issue, Au, Knightly, Jun, and Oh (2002) and Knightly, Jun, Oh, and Au (2003) studied the Spanish language skills of native English speaking college students in second year Spanish classes. Half of the participants had been exposed to Spanish without learning it or speaking it during early childhood; the other half had not. In a comparison with the speech of native Spanish speakers, the researchers found near native-like VOTs among overhearers, but significantly longer VOTs among the other Spanish learners. In addition, overhearers produced more lenition of medial voiced stops, a characteristic of Spanish, than did non-overhearers, but less than native speakers. Likewise, judgments of accent were intermediate for overhearers between judgments of native Spanish and non-overhearers' speech. On measures of morphosyntactic knowledge, overhearers and nonoverhearers were equivalent, and both showed lower performances than that of native speakers.

In short, these two studies show evidence that mere overhearing of a language that one does not speak or understand can have an impact on production of phonetic segments in the language when it is learned, but no measurable impact on learning of morphosyntactic properties of the language. We address phonetic production by comparing the VOTs of monolingual French speakers in Paris and those of monolingual English speakers in Connecticut, respectively to monolingual speakers of French and English in Montreal. We expect the latter two groups to be exposed to the language they do not know to a greater extent than the Paris and Connecticut speakers.

Relatedly, Caramazza and Yeni-Komshian (1974) compared the VOTs of monolingual French speakers in Nantes, France and Montreal, Canada. They found, for voiced consonants, a lower frequency of voicing leads and a correspondingly higher frequency of short lags in the Canadian speakers than in the speakers from

France. This difference may reflect an influence of English in which (initial) voiced stops are infrequently prevoiced. Compatibly, Canadian French speakers had longer voiceless VOTs than speakers from France, as if their voiceless stops in French were attracted toward the long-lag VOTs of voiceless stops in English. The investigators also compared the VOTs of monolingual English speakers in Montreal to published data on VOTs of monolingual English speakers not exposed to French and found no differences. They ascribed the difference in outcome in the speech of Canadian speakers of French and English to the fact that, in Montreal (at that time), “Canadian French is an island in a sea of English” (p. 244). However, another possibility is that published data on English was less comparable to that collected on Canadian English speakers than were the French data collected comparably in France and Canada. Our study provides a fairly direct follow up on that of Caramazza and Yeni-Komshian (1974) in its comparison of monolingual French and English speakers who are and are not exposed to the other language on a frequent basis.

To summarize, our study asks, in its comparison of monolingual and bilingual speakers of French and English in Montreal whether the memory systems supporting production of the two languages are independent. Our index of independence or, as we hypothesize, the lack thereof, is VOT. We hypothesize that the French VOTs of simultaneous bilinguals in Montreal will be longer than those of monolingual French speakers and that the English VOTs of simultaneous bilinguals will be shorter than those of monolingual English speakers. Second, we assess the impact on the VOTs of monolingual French and English speakers of overhearing a language they do not speak or know. If there is an effect of overhearing, we should find longer VOTs of French monolinguals in Montreal than in Paris, and shorter VOTs of English monolinguals in Montreal than in Connecticut.

1. Experiment

1.1. Method

1.1.1. Participants

Seventy-eight participants from Montreal were recruited (41 females, 37 males) to participate in the present study. The sample mainly consisted of university students recruited through notices posted on both the McGill University campus and the Université de Québec à Montréal campus, in addition to McGill online classified ads. Participants were compensated 10\$/h for their time. The mean age of the sample was 25.6 years (range 18–57).

Individuals from five categories of speakers were recruited in Montreal. Key criteria included being born and raised in Quebec or Ottawa and speaking French and/or English fluently, with no to low proficiency in other languages.

Monolingual English speakers had English as their native language, went through the English school system in Quebec or Ottawa, predominantly used English on a daily basis (85–100%) including speaking to friends and family, and rated themselves with at least a 6 or 7 on a seven-point Likert scale on their English linguistic competence, and no higher than 4 on their competence in other languages.

Monolingual French speakers had French as their native language, went through the French school system in Quebec or Ottawa, predominantly used French on a daily basis (85–100%) including speaking to friends and family, and rated themselves with at least a 6 or 7 on their French linguistic competence, and no higher than 4 on their competence in other languages.

French/English bilinguals from birth (“simultaneous” bilinguals) had both French and English as their native language, went through either the French and/or English school system in Quebec or Ottawa, used both French and English on a daily basis (at least 20% for each language) including speaking to friends and family, and rated themselves with at least a 6 or 7 on both their French and English linguistic competence, and no higher than 4 on their competence in other languages.

French/English bilinguals with English as their L1 and French as their L2 had English as their native language and learned French in primary school (typically around 4–5 years old). They went through either the French and/or English school system in Quebec or Ottawa, used both French and English on a daily basis (at least 20% for each language) including speaking to friends and family, and rated themselves with at least a 6 or 7 on

their English linguistic competence and at least a 5 on their French competence and no higher than 4 on their competence in other languages.

French/English bilinguals with French as their L1 and English as their L2 had French as their native language and learned English in primary school (typically around 9–10 years old). They went through the French school system in Quebec or Ottawa, used both French and English on a daily basis (at least 20% for each language) including speaking to friends and family, and rated themselves with at least a 6 on their French linguistic competence and at least a 5 on their English linguistic competence, and no higher than 4 on their competence in other languages.

There were 16 monolingual French speakers, 16 monolingual English speakers, 16 French L1 bilinguals, 15 English L1 bilinguals and 15 natively bilingual speakers.

Monolingual French speakers in Paris were 11 speakers (six females, five males) who were recruited through notices posted on the campus of Paris V René Descartes University. All were students at Paris V University and participated voluntarily in the study. The mean age of the sample was 22.7 years (range 19–39). All of them were raised in French monolingual families in the Île-de-France region and used French almost exclusively, including speaking to friends or watching TV (85–100%). Although they took English at school as a (mandatory) second language from the age of 11–13 years, they rated themselves with 4.05 in average (range 2.0–5.6) on their linguistic competence in English, as compared to 7 in French.

Monolingual English speakers in Mansfield, CT were 16 students at the University of Connecticut. Whereas most had taken a foreign language in school, none rated their proficiency in that language higher than 4 (mean rating 1.85 among the 12 who had taken a foreign language).

1.1.2. Stimulus materials

The stimulus materials are presented in Appendix A. In each language, they consisted of 10 sentences preceded and followed by three filler sentences. Across the 10 sentences, there were 30 target words, 10 each beginning with /p/, /t/, or /k/. The vowel contexts that followed each consonant (with the number of the syllable in which they appeared indicated) are listed after the sentences in the appendix.

1.1.3. Procedure

1.1.3.1. Montreal. Advertisements prompted potential participants to contact the experimenters via email or phone. Candidates recruited online were sent an email with a brief description of the study and were asked to either call the experimenter or to reply a few prescreening questions. Once contact was established by phone, a series of questions was asked in order to assess whether the potential subjects fit into one of the five categories of speakers the study aimed to test.

Those who had the linguistic competence and background appropriate for the study were scheduled to come to the laboratory for either a 30 min session if they were judged as monolingual or two 30 min sessions if they were considered bilingual. Bilinguals were typically scheduled to have their two sessions at least 1-day apart, although in a few cases one session was held in the morning and the other in the afternoon. Participants were told that the study involved reading sentences on a computer screen and that their speech would be recorded for later acoustic analysis.

When the English sentences were to be recorded, the task was explained in English by a bilingual experimenter. When the French sentences were to be recorded, the task was explained in French by the same experimenter. After signing the consent form, participants sat at a desk facing a computer screen with a microphone (Sennheiser) placed at a distance of about 15 cm from their mouth. The sampling rate was set at 44,1000 Hz for the majority of participants with a low pass filter at 22,000 Hz. The first six datasets were collected at a 10,000 Hz sampling rate and low pass filtered at 5000 Hz. Participants were told to read the sentences presented on the computer screen at a natural pace, speaking neither too quickly nor too slowly. They were asked to speak clearly. Sentences were presented on a computer screen one sentence at a time in a random order. In each session the sentence set was presented three times in different orders, yielding a dataset of 90 target utterances. If speakers made speech errors in producing a sentence, they repeated it before moving on.

In order to increase attendance to the second-testing session among bilinguals, the first session was typically in their native language. That is, for bilinguals with English as their L1, the first session was typically in

English and the opposite was true for bilinguals with French as their L1. For individuals with both French and English as native languages, the order was varied.

1.1.3.2. Paris. The same format of visual presentation was used in Paris. The sampling rate was set to 16,000 Hz, with low-pass filtering at 8000 Hz. The instructions were given in French.

1.1.3.3. Mansfield. Procedures in Mansfield were the same as those in Montreal except that no phone interviews were conducted. Questionnaires were administered in the same session in which students provided recordings of our stimulus materials.

1.1.4. VOT measurements

Measurements of VOT were made from waveforms using an algorithm written by Guillaume Houle and modified by Mark Tiede. Each token was displayed in a window with three panels showing the original speech signal, its corresponding broadband spectrogram, and the rectified audio with superimposed RMS. Users placed a marker to the left of visible evidence of a stop burst, and a second marker to the right of apparent voicing onset for the vowel following the voiceless stop. The algorithm used RMS amplitude to identify both the onset of the burst and the onset of voicing. We calculated RMS amplitude on a per sample basis using a moving rectangular window 4 ms in length. Burst onset was first sample that preceded voice onset by 8 ms or more and had a magnitude that exceeded 40% of the RMS maximum. Voice onset was defined as the first sample that exceeded 50% of the maximum RMS for that token. On the infrequent occasions on which the algorithm made an obvious error, users could make corrections in marker placement by hand. Because the vast majority of measures were based on automatic extraction of VOT by the algorithm, which would give the same VOT measures to all users, we did not collect reliability measures across measurers. However, for measures of the Montreal speakers, we did assess the measurement error associated with the algorithm, by comparing measures that included or excluded hand corrections. (Given the outcome, we did not repeat this comparison for Paris and Mansfield speakers.) That is, we compared VOTs obtained only automatically with VOTs that included hand corrections. Across speakers, the average measurement difference was less than 1 ms.

2. Results¹

We first focused on our monolingual and simultaneous bilingual speakers of French and English in Montreal to ask whether VOTs in the two languages of simultaneous bilingual speakers were influenced by their other language. Fig. 1 shows the data separately for /p/, /t/ and /k/. Each plot shows mean VOTs (± 1 SEM) in French and English for monolingual (light bars) and bilingual speakers. In a three-way ANOVA with the within-subjects factor consonant and between-subjects factors language² (French, English) and group (monolingual, bilingual), the effect of consonant was highly significant ($F(2, 116) = 511.33, p < .001$) reflecting the expected finding that VOT increases as place of articulation moves back in the vocal tract. Fig. 1 also shows that the same pattern of VOTs is observed for all three consonants ($F < 1$ for the three-way interaction of consonant by language by group). Accordingly, we collapsed over consonants and repeated our analyses in order to focus on differences between languages and monolingual and bilingual speaker groups. The effect of language was highly significant ($F(1, 182) = 411.41, p < .001$) reflecting the also expected finding that VOTs of French voiceless unaspirated stops are shorter than English voiceless aspirated stops. The effect of group was not significant ($F < 1$). However, the predicted group by language interaction did reach significance

¹A reviewer expressed an interest in any sex differences that we might have seen in our data. Because an examination of sex differences was not a goal of our research, and presentation of those findings in the results would deflect attention from the main purposes of the research, we present findings of those analyses in Appendix B. In short, we obtained no findings of interest. In particular, we find no evidence for the finding sometimes obtained (see Whiteside, Henry, & Dobbin, 2004, for a review) that females have longer VOTs than males. More relevantly to our purposes, we find no evidence that females and males differ as a function of bilingual or monolingual language group in their VOTs in French and/or English.

²Note that this factor is a within subjects factor for the bilingual group, but a between subjects factor for the monolingual group.

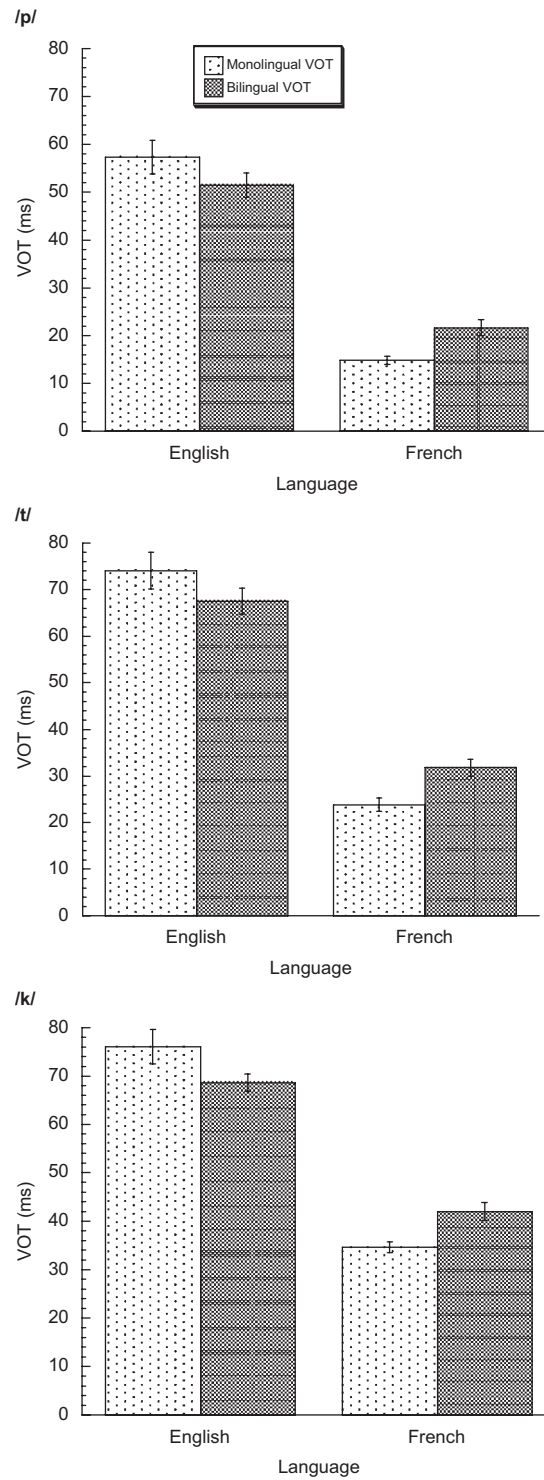


Fig. 1. Voice onset times (± 1 SEM) of simultaneous bilingual speakers and monolingual speakers of English and French from Montreal. Consonants are /p/ (top), /t/ (middle) and /k/ (bottom).

($F(1, 182) = 14.19, p < .001$), with language differences in VOT being smaller for the bilingual speakers than for the monolingual speakers.

In Bonferroni-corrected post-hoc comparisons (collapsed across the three consonants), the 6.6 ms difference in English monolingual vs bilingual VOTs was significant, $p = .013$; the -7.4 ms difference between the French speaking groups also reached significance ($p = .005$).

One additional two-way interaction (consonant by language) from the original three-way analysis reached significance ($F(2, 116) = 29.64, p < .001$) probably because the difference in VOT between /t/ and /k/ was much smaller in English (2 ms) than in French (10 ms).

We next looked at all groups of French monolingual and bilingual speakers from Montreal. Fig. 2 shows the data. As expected, VOTs are shortest for /p/ and longest for /k/. In an analysis of variance with factors consonant and speaker group, the effect of consonant was highly significant ($F(2, 116) = 643.20, p < .001$). There was also an effect of speaker group ($F(3, 58) = 14.60, p < .001$) with monolingual speakers producing the shortest VOTs and English L1 bilingual speakers producing the longest VOTs. In Tukey post-hoc comparisons, aside from the difference between monolinguals and simultaneous bilinguals already described, English L1 bilinguals differed both from French L1 bilinguals and from monolingual speakers (both $ps < .001$). The interaction of consonant and speaker group did not approach significance ($F < 1$).

Fig. 3 provides the complementary findings for groups of Montreal speakers who were speaking English. As for the French VOTs, those for English /p/ were shortest, and those for /k/ generally longest. In an analysis of variance with factors consonant and speaker group, the effect of consonant was highly significant ($F(2, 116) = 360.63, p < .001$). The effect of group was also significant ($F(3, 58) = 8.17, p < .001$) with monolingual English speakers showing the longest VOTs and French L1 bilinguals the shortest. In Tukey post-hoc comparisons, the French L1 bilinguals differed from all other groups (largest $p < .02$). In this case, the interaction of consonant and speaker group did reach significance ($F(6, 116) = 4.36, p = .001$), reflecting the fact that English L1 bilinguals had shorter VOTs for /k/ than /t/ whereas the other groups showed the opposite, and expected, direction of difference.

Another question we can ask about our bilingual speakers is whether they produce French and English voiceless stops with different VOTs. That is, if their VOTs in both languages differ from those of monolinguals in analogous ways, is it because they have only one way of producing /p/, one for /t/ and one for /k/ that serves both languages? In a single overall analysis with factors language and consonant, all three bilingual groups

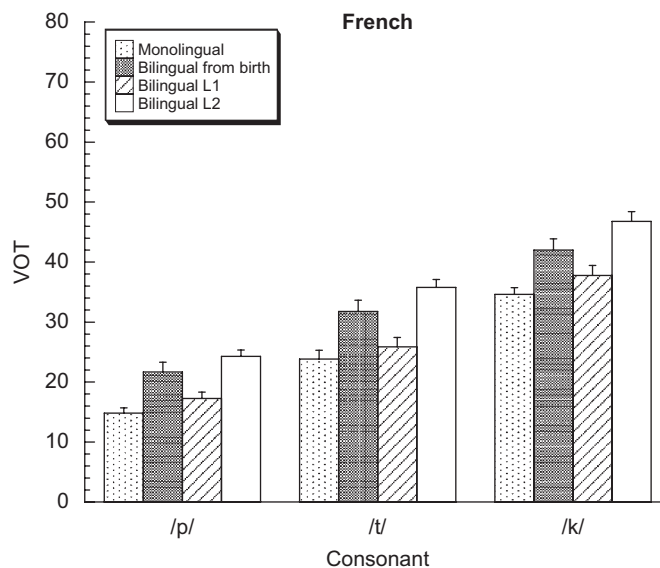


Fig. 2. Voice onset times (± 1 SEM) of Montreal speakers speaking French, with standard errors indicated. Data are from monolingual speakers of French, speakers who are bilingual from birth, bilingual speakers who are speaking their first learned language (Bilingual L1), and bilingual speakers speaking their second learned language (Bilingual L2).

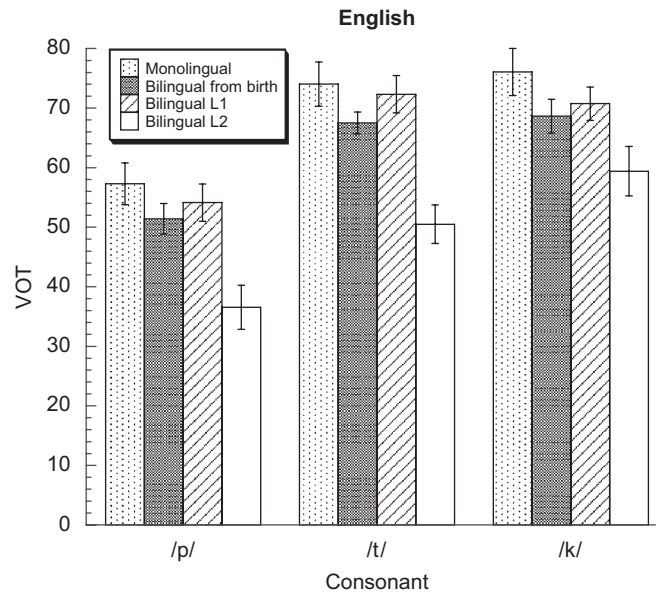


Fig. 3. Voice onset times (± 1 SEM) of Montreal speakers speaking English, with standard errors indicated. Data are from monolingual speakers of English, speakers who are bilingual from birth, bilingual speakers who are speaking their first learned language (Bilingual L1), and bilingual speakers speaking their second learned language (Bilingual L2).

showed highly significant effects of language on VOTs for the same consonant. In every numerical comparison VOTs were longer in the bilinguals' English than in their French consonants, and all pairwise within-consonant comparisons were highly significant by Bonferroni post-hoc comparisons ($p < .0001$ for all speaker groups and consonants).

To assess effects of overhearing a language otherwise unknown to speakers, we compared the VOTs of Montreal and Parisian French monolinguals and Montreal and Mansfield, CT English monolinguals. In our data, there were no such effects. On average monolingual French speakers in Montreal had VOTs of 24.4 ms; those of the Parisians averaged 26.3 ms, a nonsignificant difference in the wrong direction for the hypothesis. Montreal English monolinguals' VOTs averaged 69.1 ms; those of Connecticut speakers averaged 70.4 ms, another nonsignificant difference.

3. Discussion

Our study was designed to ask whether natively bilingual speakers produce speech in each language that, in a sense, is accented due to an influence from the other language. This might occur, according to Flege's Speech Learning Model (e.g., Flege, 1995), because corresponding members of bilingual language users' sound inventories, in this case, the voiceless stops, are cognitively identified with one another. Our data answer that question in the affirmative. Among French speakers, monolingual speakers produce VOTs that are significantly shorter than those of natively bilingual speakers (and natively bilingual speakers have VOTs that are marginally longer than those of French L1 bilinguals, who should be less influenced by English). The numerical pattern was analogous in English. That is, monolingual English speakers had the longest VOTs, those of English L1 bilinguals were next longest and longer than those of natively bilingual speakers. The difference between the monolinguals' and simultaneous bilinguals' VOTs was marginally significant.

The analogous outcomes in the two languages do not occur because our bilingual speakers produced voiceless stops identically in the two languages. All three groups of bilinguals have distinct voiceless stops in English and French. This finding of a link between, but not a merging of, phones that are phonetically distinct but similar in the two languages of a bilingual speaker is consistent with the concept of "equivalence classification" in Flege's Speech Learning Model as described in the introduction. That is, bilingual speakers treat the phones of L2 that are sufficiently similar to phones in L1 as variants of the L1 phones.

This is the same outcome that [Sancier and Fowler \(1997\)](#) found for their Portuguese speaker. The outcome for the Portuguese speaker is perhaps more striking than the present one, because her English and Portuguese speech both changed when the speaker was exposed to just one of the languages (as it might be expected to as a consequence of equivalence classification). That is, her English VOTs shortened when she was in an environment in which essentially only Portuguese was being spoken. Similarly, her Portuguese VOTs lengthened in an environment when essentially only English was being spoken.

Our outcome on French VOTs is different from that of [Sundara et al. \(2006\)](#), who found neither numerical nor statistical differences between French simultaneous bilinguals and French monolinguals in production of /t/. Our findings on /t/ paralleled those on /p/ and /k/ in showing significantly shorter VOTs for monolingual speakers. This difference may reflect our larger sample size. It is not obvious that any differences in criteria for identifying simultaneous bilinguals in the two studies could explain the differences in outcome.

Comparisons of monolingual French and English speakers in Montreal with Parisian French and Mansfield, CT English speakers were meant to address the question whether any cross language influences we might find reflected mere exposure to a second language or whether active use of the language was required. Under the assumption that Montreal monolingual French speakers hear more English than do Parisian French monolinguals and under the very likely assumption that monolingual speakers of English in Montreal hear more French than do Connecticut monolinguals, our findings permit a clear interpretation. VOTs were nearly identical for the monolingual French speakers in the two locations; they were also closely similar for the monolingual English groups.

This outcome is different from that of [Caramazza and Yeni-Komshian \(1974\)](#), who found a VOT difference between monolingual French speakers in Montreal and Nantes, France, but no difference between Canadian speakers of English and US English speakers from published data. They ascribed the asymmetry in outcome to the fact that Canadian French speakers were islands in a sea of English speech. This may be less so today than it was 30 years ago. Their outcome with English speakers was similar to ours. However, our null result may be more secure, because we tested participants in Connecticut and Quebec with the same materials whereas they compared Canadian English speaker in their sample with published data.

Although [Au et al. \(2002\)](#) and [Knightly et al. \(2003\)](#) also found effects of overhearing on speech production, their findings were on learning the phonological categories of an L2 and so are not directly relevant to the findings of the present study.

Our findings help to strengthen evidence that cross language phonetic influences do occur in the speech of bilingual speakers. However, they do not help to establish why they occur, except in suggesting that the influences occur only or primarily when a speaker actively uses another language. Understanding would be improved if we knew more about the nature of the influences that occur. For example, is the influence, in fact, restricted to phonetic segments that language users detect as corresponding in the two languages, as the Speech Learning Model suggests? It is intuitively unlikely that a bilingual speaker of a click language, such as Zulu, and a language without clicks, such as English or French, would be influenced either *in* click production or *by* click production in uttering nonclick sounds in the other language. However, imagine a speaker of Hawaiian and of English. Hawaiian has unaspirated /p/ and /k/, but no unaspirated alveolar or dental stop ([Maddieson, 1984](#)). English has (stressed syllable initial) aspirated voiceless stops that include /t/, an alveolar. On the basis of the present findings, we would expect the Hawaiian /p/s and /k/s of the hypothetical Hawaiian bilingual to be more aspirated (to have longer VOTs) than those of a monolingual Hawaiian speaker. And we would expect the English /p/s and /k/s to have shorter VOTs than those of monolingual English speakers. But what about English /t/? It does not correspond with any Hawaiian consonant. So one possibility is that the speaker's VOTs of /t/ would not be shorter than those of a monolingual speaker. However, /t/ belongs to a system of aspirated voiceless stops. If VOTs of /p/ and /k/ are affected by speakers knowing a language with unaspirated /p/s and /k/s, perhaps those of /t/ would be as well.

A related question is, if speaking a second language leads to phonetic influences on a first language, why do the segments of a given language not cause other segments in that language to drift? (Or, perhaps the question is *whether* the segments of a language cause one another to drift.) In Hindi, for example, aspirated and unaspirated voiceless stops are distinct phonemes. Why is each kind of stop not influenced by the other? If that is the right question (as opposed to: Is each kind of stop influenced by the other?), one answer might again resort to equivalence classification. For speakers of Hindi, aspirated and unaspirated /p/, for example, are not

identified as variants of the same consonant (in the same way that English aspirated and unaspirated /p/ are so identified by speakers of English). Perhaps an influence is only exerted by segments identified as variants of the same consonant or vowel. If the better question is whether the segments of a given language can influence one another, we might ask whether monolingual speakers of Hindi have longer unaspirated stop VOTs and shorter aspirated stop VOTs than monolingual speakers of languages that, respectively, only have unaspirated or only aspirated voiceless stops. Do aspirated stops in English have shorter VOTs, because of their correspondence with unaspirated stops that occur in other positions in a word, shorter than in languages that only have aspirated stops?

A final question to raise here is what the dimensions are along which cross language phonetic influences occur. There has been a fairly intensive (albeit not exclusive) interest in VOT, perhaps because it is easy to measure acoustically. Does an influence occur on place of articulation? For example, do speakers of two languages, one of which has alveolar and one dental stops, produce the alveolar stops with a more forward place of articulation and the dental stops with a more back place of articulation than monolingual speakers of each language? Do the vowel inventories of bilingual speakers' two languages show shifts in height or fronting under influence from the other language (cf. Guion, 2003)?

We leave these and other interesting questions for future research.

Acknowledgments

This research was supported by NIDCD grants DC-03782, DC-02717, and DC-04669 to Haskins Laboratories. We thank Guillaume Houle and Mark Tiede for generating and revising, respectively, the algorithm that we used to measure VOT.

Appendix A

A.1. English sentences

1. Ben bought some flowers and put them on his dining room table.
2. Seven hungry children crowded around the buffet.
3. Miranda's job was boring, and she fell asleep at her desk.
4. At the store, **Kate purchased a tape** recorder and a new stereo.
/kei/ (s4), /per/ (s5), /tei/ (s8)
5. As recently as **two** days ago, Lucy **parked** her **car** at the grocery store, and she forgot where she left it.
/tu/ (s2), /pa/ (s6), /ka/ (s8)
6. Fred wore a heavy **parka** and **comfortable** boots on the hike up **Tabletop** Mountain.
/pa/ (s6), /ko/ (s9), /te/ (s18)
7. Driving along the **turnpike**, **Kayla** listened to **polkas** on the radio.
/ter/ (s6), /kei/ (s8), /po/ (s13)
8. On his **perch**, the **tiny** bird **called** to his mate.
/per/ (s3), /tai/ (s5), /ko/ (s8)
9. Braving the raging surf, **Peter caught a towering** wave and rode his surfboard to shore.
/pi/ (s7), /ko/ (s9), /tau/ (s11)
10. Over the holiday weekend, Marvin **performed** his magic tricks, **keeping** his brother **Tommy** amazed and amused.
/per/ (s11), /ki/ (s17), /to/ (s22)
11. **Bonnie covered** the stewed **tomatoes** and turned down the burner before starting to work on some **pies** for dessert.
/ko/ (s3), /to/ (s7), /pai/ (s24)
12. Every **time** he sneaked down the stairs hoping to get himself a snack, **Paul's** wife caught him and handed him a **carrot** or a piece of celery.
/tai/ (s3), /po/ (s17), /ka/ (s26)

13. Depressed that the dentist had found three **cavities**, **Tim pestered** his mother to buy him some chocolate candy.
/ka/ (s10), /ti/ (s13), /pe/ (s14)
14. While waiting for his car to be fixed, Linda watched TV.
15. Looking through the telescope, the students saw Venus.
16. Colin browsed in the bookstore while his sister shopped for a new briefcase.
/kei/ (s4), /per/ (s5), /tei/ (s8)
/tu/ (s2), /pa/ (s6), /ka/ (s8)
/pa/ (s6), /ko/ (s9), /te/ (s13)
/ter/ (s6), /kei/ (s8), /po/ (s13)
/per/ (s3), /tai/ (s5), /ko/ (s8)
/pi/ (s7), /ko/ (s9), /tau/ (s11)
/pcr/ (s11), /ki/ (s17), /to/ (s22)
/ko/ (s3), /to/ (s7), /pai/ (s24)
/tai/ (s3), /po/ (s17), /ka/ (s26)
/ka/ (s10), /ti/ (s13), /pe/ (s14)

A.2. French sentences

1. Il a acheté des roses et les a mises dans un très beau vase.
2. Les enfants, affamés, prenaient d'assaut le buffet.
3. Son travail est si ennuyeux qu'elle s'est endormie sur son bureau.
4. Avant-hier, **Catherine** a eu **peur** des **termites** qui ont dévoré les poutres.
/ka/ (s4), /peu/ (s8), /tè/ (s10)
5. C'est **tout** ce que l'ami **Paul** avait **caché** dans ses affaires: une vieille boîte d'allumettes.
/tu/ (s2), /po/ (s6), /ka/ (s9)
6. Dans la doublure de sa **parka**, il **cachait** au moins une bonne centaine de **timbres** de Bosnie-Herzégovine.
/pa/ (s7), /ka/ (s10), /tè/ (s19)
7. A l'autre bout du **terrain**, **Karine** dansait la **polka** avec Ronaldo.
/té/ (s6), /ka/ (s8), /po/ (s13)
8. J'ai eu **peur**: ce **taxi** me **collait** de trop près.
/pœ/ (s3), /ta/ (s5), /ko/ (s8)
9. Lorsque j'ai visité **Pékin**, un **cortège** de **taoïstes** défilait devant l'ancienne cité impériale.
/pé/ (s7), /ko/ (s9), /tao/ (s13)
10. C'est sûrement au cours de la soirée qu'il a **perdu** le précieux **kimono** de son frère **Tobie**, un cadeau de leur grand-mère.
/pè/ (s12), /ki/ (s17), /to/ (s22)
11. Dans le **couvent** ce sont les **tomates** qui semblent être l'objet de tous les soins; les nonnes s'inquiètent de leur **pâleur** cette année.
/ku/ (s3), /to/ (s8), /pa/ (s25)
12. Le grand **tapis** iranien aux motifs si sophistiqués, **posé** là en attendant, donnait à **Caroline** un peu de répit.
/ta/ (s3), /pa/ (s16), /ka/ (s26)
13. Comme le dentiste lui avait trouvé des **caries**, **Tim pestait** contre sa sœur qui lui donnait trop de bonbons.
/ka/ (s11), /ti/ (s13), /pe/ (s14)
14. En attendant que le café soit prêt, Laure lisait le journal.
15. Grâce au nouveau télescope, on peut voir **Vénus**.
16. Nicolas fouinait dans les rayons de la librairie pour le plaisir.
/ka/ (s4), /peu/ (s8), /tè/ (s10)
/tu/ (s2), /po/ (s6), /ka/ (s9)
/pa/ (s7), /ka/ (s10), /tè/ (s19)

/té/ (s6), /ka/ (s8), /po/ (13)
 /pœ/ (s3), /ta/ (s5), /ko/ (s8)
 /pé/ (s7), /ko/ (s9), /tao/ (s13)
 /pè/ (s12), /ki/ (s17), /to/ (s22)
 /ku/ (s3), /to/ (s8), /pa/ (s25)
 /ta/ (s3), /pa/ (s16), /ka/ (s26)
 /ka/ (s11), /ti/ (s13), /pe/ (s13)

Appendix B

We did not design our research intending to look for sex differences; accordingly, we made no effort to balance the numbers of men and women in each group. In the group of participants who are bilingual from birth, there were just 3 men and 12 women. Among other French speakers, there were 6 men and 10 women in the French L1 group, 9 women and 6 men in the English L1 group, 11 men and 5 women in the monolingual group, and 6 men and 5 women from the Parisian French monolingual group. Among English speakers, there were 6 men and 10 women in the French L1 group, 6 men and 9 women in the English L1 group, 10 men and 6 women in the monolingual speakers from Montreal, and 5 men and 11 women among the monolingual speakers from Mansfield, CT.

To explore any sex differences in the patterning of VOTs among our groups of participants, we performed four ANOVAs, two on data from Montreal speakers of French and two on data from Montreal speakers of English. One analysis in each language had factors speaker group, with levels: Bilingual from birth, Bilingual-L1 French, Bilingual L1 English, and Monolingual (Montreal speakers), sex, and consonant. The other two analyses compared monolingual speakers from Montreal with monolingual speakers from Paris (French) or Connecticut (English). In the following, we discuss only outcomes involving the factor sex, because other outcomes are addressed in Section 2.

In the analysis of bilingual and monolingual Montreal speakers of French, the main effect of sex did not approach significance ($F(1, 54) = 1.18, p = .163$), and no interactions involving the sex factor were significant. However, the interaction of sex by consonant approached significance ($F(2, 108) = 3.02, p = .053$), apparently because female VOTs on /p/ exceeded male VOTs (by 1 ms) whereas they were shorter (also by 1 ms in each case) for /t/ and /k/.

In the analysis of bilingual and monolingual Montreal speakers of English, the outcome was similar. The main effect of sex did not approach significance ($F < 1$), and it did not participate in significant interactions with other factors (all $F_s < 1$).

In the analysis of the French monolingual groups in Montreal and Paris, the effect of sex did not approach significance ($F(1, 23) = 1.37, p = .25$), and no interactions involving the sex factor approached significance (smallest p value: .32).

In the analysis of the English monolingual groups in Montreal and Mansfield, the main effect of sex did not approach significance ($F < 1$). Just one interaction, sex by consonant by group, reached significance. This was because the Mansfield women showed no difference in VOT between /t/ and /k/ whereas all other groups showed the expected longer VOT for /k/.

References

- Au, T. K., Knightly, L. M., Jun, S.-A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science, 13*, 238–243.
- Bourhis, R., & Giles, H. (1977). The language of intergroup distinctiveness. In H. Giles (Ed.), *Language, ethnicity and intergroup relations* (pp. 119–135). London: Academic Press.
- Caramazza, A., & Yeni-Komshian, G. H. (1974). Voice onset time in two French dialects. *Journal of Phonetics, 2*, 239–245.
- Clark, H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Condon, W. (1976). An analysis of behavioral organization. *Sign Language Studies, 13*, 285–318.
- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47–65.
- 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. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, 6, 185–202.
- Giles, H. (1973). Accent mobility: Models and some data. *Anthropological Linguistics*, 15, 87–105.
- Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge: Cambridge University Press.
- Guion, S. (2003). The vowel systems of Quichua-Spanish bilinguals. *Phonetica*, 60, 98–128.
- Jaffe, J., & Feldstein, S. (1970). *Rhythms of dialogue*. New York: Academic Press.
- Knightly, L. M., Jun, S.-A., Oh, J. S., & Au, T. K. (2003). Production benefits of childhood overhearing. *Journal of the Acoustical Society of America*, 114, 465–474.
- Labov, W. (1963). The social motivation of a sound change. *Word*, 19, 273–309.
- LaFrance, M. (1982). Posture matching and rapport. In M. Davis (Ed.), *Interaction rhythms: Periodicity in Communicative Behavior* (pp. 279–297). New York: Human Sciences Press.
- Mack, M. (1989). Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Perception & Psychophysics*, 46, 187–200.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.
- Natale, M. (1975). Convergence of mean vocal intensity in dyadic communications as a function of social desirability. *Journal of Personality and Social Psychology*, 32, 790–804.
- Sancier, M., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25, 421–436.
- Shockley, K., Santana, M.-V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in the supra-postural task of cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 326–332.
- Street, R. L. (1983). Noncontent speech convergence in adult-child interactions. In R. N. Bostrom (Ed.), *Communication yearbook 7* (pp. 369–395). Beverly Hills, CA: Sage.
- Sundara, M., Polka, L., & Baum, S. (2006). Production of coronal stops by simultaneously bilingual adults. *Bilingualism: Language and Cognition*, 9, 97–114.
- Whiteside, S., Henry, L., & Dobbin, R. (2004). Sex differences in voice onset time: A developmental study of phonetic context effects in British English. *Journal of the Acoustical Society of America*, 116, 1179–1183.