Consonantal voicing effects on vowel duration in Italian-English bilinguals

Ylana Beller-Marino

Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: http://academicworks.cuny.edu/gc_etds

Part of the Linguistics Commons

Recommended Citation

Beller-Marino, Ylana, "Consonantal voicing effects on vowel duration in Italian-English bilinguals" (2014). CUNY Academic Works.
http://academicworks.cuny.edu/gc_etds/339

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Graduate Works by Year: Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
CONSONANTAL VOICING EFFECTS ON VOWEL DURATION
IN ITALIAN-ENGLISH BILINGUALS

by

YLANA BELLER-MARINO

A dissertation submitted to the Graduate Faculty in Linguistics in partial fulfillment of the
requirements for the degree of Doctor of Philosophy, The City University of New York

2014
This manuscript has been read and accepted for the Graduate Faculty in Linguistics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Dr. Dianne Bradley

Date

Chair of Examining Committee

Dr. Gita Martohardjono

Date

Executive Officer

Dr. Eva Fernández

Dr. Douglas H. Whalen

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
Abstract

CONSONANTAL VOICING EFFECTS ON VOWEL DURATION IN ITALIAN-ENGLISH BILINGUALS

by

Ylana Beller-Marino

Advisor: Professor Dianne Bradley

This project reported in this dissertation analyzes phonetic details of the speech patterns in one of New York’s bilingual communities, asking whether a bilingual speaker can attain native-like proficiency in both languages and the extent to which authenticity — maintenance of language-specific settings — is sustainable. Researchers have established that Italian and English differ strikingly in their characteristic time settings for vowel durations: durations are greater for vowels preceding voiced consonants, e.g., cab, rather than voiceless, e.g., cap. This duration difference, termed the consonantal voicing effect (CVE), is notably greater for English than for Italian. The greater magnitude of the CVE found with English is considered to be a phonological enhancement of a basic phonetic process. Utilizing a speech production task, the study reported compares the performance of Italian-born bilinguals for whom English was acquired in adulthood, as a second language, with that of U.S.-born speakers who experienced simultaneous acquisition of their languages (albeit in an English-dominant setting).

In separate sessions for each language, speakers produced utterances in which the target word, situated inside a carrier phrase, contrasted in [voice] value for the post-vocalic consonant, e.g., Say the word « ___ » to me. Stimuli were familiar words selected to sample the vowel
inventories for each language and for which the voicing contrast was realized through the
inventory of stops common to both languages. Analyses revealed no evidence of influence of the
second language on the CVE for the first language for either group, despite an extended
immersion period in an English-language environment for the foreign-born speakers and
simultaneous exposure to both languages from birth for the U.S.-born speakers. But crucially,
there was evidence of an influence of the first language in the timing settings found for the CVE
in the second language, for both speaker groups: the foreign-born speakers managed to increase
the magnitude of the CVE-English but failed to fully implement the phonological mechanism
consistent with larger CVE values for that language; and the U.S.-born speakers managed to
reduce the magnitude of the CVE-Italian but failed to fully suppress that same mechanism.
Results are discussed in relation to language-specific timing patterns and the extent to which a
dominant language may influence production in the non-dominant language.
Acknowledgments

I wish to first thank my dissertation advisor, Dr. Dianne Bradley, for her endless support and patience throughout each and every stage of this project. I went to Dr. Bradley with an idea for a project, and she helped me turn the idea into a reality. Under her guidance, I grew and developed my research skills immeasurably. She ensured that I gave ample consideration to the small yet critical details of the project while keeping me focused on the bigger picture. Her mastery of stylistics is unparalleled. I will not soon forget, “a datum has a value; a girl has a figure.” Dr. Bradley has been incredibly generous with her time, her knowledge and her wisdom. I must also acknowledge that she expressed extreme kindness and compassion as I traversed numerous personal events, good and otherwise, during the course of my graduate school career.

I would also like to thank my committee members, Dr. Eva Fernández and Dr. Doug Whalen. Dr. Fernández has been incredibly supportive, always sharing her expertise, asking thoughtful questions and offering insightful suggestions. In addition to being my technology guru, she has given me invaluable practical advice on the work-life balance as a graduate student. Dr. Whalen has been an extremely valuable addition to my committee. I benefitted from his vast knowledge and rational approach. Being part of Dr. Whalen’s lab additionally gave me a family at school that I had previously lacked.

I am grateful to the members of the Quagliettana Society in Queens, who not only provided partial inspiration for my project, but also served as participants in the study. The residents of Quaglietta were welcoming and kind during my time in Italy. In particular, I wish to thank zia Assunta and the Russo family, Mimina, Nicola and Luca, for their hospitality and for helping me get better acquainted with the community; and Carmelina Rizzo, for numerous
thoughtful conversations about language and life in Quaglietta and abroad. I am truly
appreciative of Michelle Johnson for her assistance in the initial segmentation of much of the
English-language data, which helped me get over what at the time seemed to be an
insurmountable hurdle. And a special thank you to both Marisa Monteleone and JungMoon
Hyun, who have both been incredibly kind and supportive during the writing stage.

I am also thankful for the many learning opportunities provided to me by the Linguistics
Program of the CUNY Graduate Center, both inside and outside the classroom. This growth has
come from my interaction with a number of professors including Dr. Ricardo Otheguy and Dr.
Gita Martohardjono, from my time at RISLUS; Dr. Martin Chodorow, who provided wonderful
direction through much of my coursework in computational linguistics; Dr. Janet Fodor, under
whose guidance I helped launch a professional development series for students within the
linguistics program; and Dr. Robert Vago, for comments on my project in its early stages and for
numerous delightful conversations related to linguistics.

I also wish to mention the people who have served as unofficial mentors to me, whether
in my undergraduate or graduate career, and whose influence led me to the present stage. These
include Dr. Carol Genetti, who ignited my passion for linguistics while I was an undergraduate;
Dr. Bernadette Luciano; Dr. Grazia Busà; and Dr. Dalton Conley, who showed me first-hand that
being a successful academic and being a good parent are not mutually exclusive. I have had the
good fortune of having numerous friends and family — including my mother, Sheila Beller, who
has been especially supportive in the last year of this endeavor, sister, brother and extended
family — at my side throughout this experience.

This work would not have been possible without my husband, Carmine Marino, who has
been supportive throughout my numerous years of graduate school and always believed I would
bring this project to fruition. He and our two daughters, Sabina and Carmina, have brought me
countless hours of happiness. Their love and devotion have been both motivational and a source
of strength.

This dissertation is dedicated to my late father,

Dr. Leonard B. Beller, who always believed I could do anything.
Table of Contents

Abstract .................................................................................................................................... iv
Acknowledgments ..................................................................................................................... vi
Table of Contents ...................................................................................................................... ix
List of Tables ........................................................................................................................... xii
List of Figures ..........................................................................................................................xiv
Chapter 1:  Introduction ...........................................................................................................1
  1.1  Vowel duration .............................................................................................................3
  1.1.1  Consonantal voicing effect in English .................................................................3
  1.1.2  Consonantal voicing effect cross-linguistically ....................................................9
  1.1.3  Duration as a cue .................................................................................................14
  1.1.4  Motivation for variation in vowel duration .........................................................16
  1.2  Second language acquisition / bilingualism ..............................................................21
  1.2.1  Age effects in acquisition ..................................................................................22
  1.2.2  Pronunciation as a function of age ....................................................................24
  1.3  Models of second language speech perception and production ............................27
  1.4  Bilingual production .................................................................................................30
  1.4.1  Inter-speech posture .........................................................................................31
  1.4.2  Voice onset time ...............................................................................................34
  1.4.3  Vowel duration in the second language ............................................................37
  1.4.4  The present study ..............................................................................................38
Chapter 2:  Implementation of the experiment .................................................................42
  2.1  Project design considerations ..................................................................................42
2.2 Project implementation detail.................................................................................................48
2.2.1 Participants: Foreign-born and U.S.-born ........................................................................49
2.2.2 Materials design................................................................................................................50
2.2.3 Elicitation protocol............................................................................................................51
2.2.4 Utterance segmentation.....................................................................................................53
2.2.5 Data extraction and filtering..............................................................................................56
2.2.6 Considerations shaping the data analysis .........................................................................60
Chapter 3: Data patterns and discussion .................................................................................62
3.1 Consonantal voicing effect: Core cases ..................................................................................62
3.1.1 Speaker speech rates .......................................................................................................68
3.1.2 Intrinsic vowel duration ..................................................................................................73
3.2 Consonantal voicing effect: The heterosyllabic environment ..............................................74
3.3 Comparing the core and heterosyllabic cases ......................................................................78
3.3.1 Syllabicity effects ..........................................................................................................80
3.3.2 More general timing patterns ........................................................................................81
3.4 Consonantal voicing effect: Flapping ...................................................................................83
3.5 Other regions .......................................................................................................................87
3.5.1 Closure: Core cases .......................................................................................................88
3.5.2 Closure: The heterosyllabic environment .....................................................................90
3.6 Summary of findings ............................................................................................................93
Chapter 4: Concluding remarks ...............................................................................................95
4.1 Consonantal voicing effects in dominant and non-dominant languages...............................95
4.1.1 Core data: Italian multisyllabic words, English monosyllabic words............................96
4.1.2 Additional data: Multisyllabic words.................................................................97
4.1.3 Data for the case of flapping .............................................................................99
4.2 Alternative approaches with the current study ......................................................100
4.3 Directions for future research .............................................................................101
Appendices ...............................................................................................................104
References ..............................................................................................................121
# List of Tables

| Table 1.1 | Vowel duration values in milliseconds in the voiced and voiceless environments, as a function of manner of articulation as reported for English by Peterson and Lehiste (1960). | 5 |
| Table 1.2 | Average vowel duration, in milliseconds, by consonantal environment, where [-VD] = voiceless stop and [+VD] = voiced stop, as reported by Hillenbrand et al. (2001). | 8 |
| Table 1.3 | Voiced to voiceless ratios for English, with corresponding study. | 9 |
| Table 1.4 | Vowel duration measurements, in milliseconds, as a function of [voice] value in the post-vocalic environment, for a single word pair in each of the four languages as reported by Chen (1970). | 12 |
| Table 1.5 | Voiced to voiceless ratios for languages other than English, with corresponding study. | 13 |
| Table 1.6 | Voice onset time values for stops in Spanish and English (from Zampini & Green 2001). | 34 |
| Table 1.7 | Mean voice onset time values for three speaker groups. | 36 |
| Table 2.1 | Illustrative stimulus word pairs, categorized by target vowel height and place of articulation of CVE-triggering consonant. | 50 |
| Table 2.2 | Mean number of vowel-duration measurements deemed usable, per participant, as a function of stimulus type. Percentages cast values for usability relative to the maximum possible. | 57 |
| Table 2.3 | Mean number of vowel-duration outliers, per participant, as a function of stimulus type. Percentages indicate values based on usable data. | 60 |
| Table 3.1 | Consonantal voicing effect (and standard error) in milliseconds for foreign- and U.S.-born speaker groups for the core cases, as a function of language of utterance. | 64 |
| Table 3.2 | Summary of hypothesized CVE mechanisms for each speaker group, as a function of language of utterance. | 65 |
| Table 3.3 | Voiced-to-voiceless ratios calculated for foreign- and U.S.-born speaker groups for the core cases, based on language of utterance. | 67 |
| Table 3.4 | Mean duration (and standard error) in milliseconds of rate estimator region for foreign- and U.S.-born speakers, as a function of language of utterance (values reported are subject-based). | 70 |
| Table 3.5 | Consonantal voicing effect (and standard error) in milliseconds for foreign- and U.S.-born speaker groups for the multisyllabic English and a subset of the Italian stimuli, as a function of language of utterance. .......................76 |
| Table 3.6 | Voiced-to-voiceless ratios calculated for foreign- and U.S.-born speaker groups for the multisyllabic English stimuli and a subset of the Italian stimuli, based on language of utterance. .............................................................77 |
| Table 3.7 | Vowel duration means in milliseconds (and percent, with the voiced monosyllable serving as the default duration) for monosyllabic and multisyllabic words in English, as a function of [voice] value in the post-vocalic environment. ..................................................................................81 |
| Table 3.8 | Consonantal voicing effect (and standard error) in milliseconds for U.S.-born speakers as a function of place of articulation of the trigger consonant. .........................................................................................................................86 |
List of Figures

Figure 1.1 Voiced to voiceless ratios for English, across studies. ........................................9

Figure 1.2 Voiced to voiceless ratios, across studies for English and other languages. The lighter markers represent values for languages other than English, while the darker markers represent values for English. ........................................14

Figure 2.1 Map displaying Italian regions. The region of Campania is circled, and the location of Quaglietta is denoted by a star. .........................................................44

Figure 2.2 Illustration of spectrogram-supported waveform segmentation and transcription for the utterance, *Dico la parola « tipo » di nuovo*. ........................54

Figure 3.1 Mean duration (and standard error) of critical vowel for foreign- and U.S.-born speaker groups for the core cases, as a function of language of utterance and [voice] value in the post-vocalic environment........................................63

Figure 3.2 Mean duration (and standard error) of rate estimator region for foreign- and U.S.-born speaker groups, as a function of language of utterance and [voice] value in the post-vocalic environment. ........................................69

Figure 3.3 Mean duration of the initial portion of the carrier phrase, the rate estimator, for each speaker, as a function of language of utterance. The triangular marker sets apart an especially slow speaker within the foreign-born group....71

Figure 3.4 Mean duration (and standard error) of critical vowel for the multisyllabic English and a subset of the Italian stimuli for foreign- and U.S.-born speaker groups, as a function of language of utterance and [voice] value in the post-vocalic environment........................................75

Figure 3.5 Mean duration (and standard error) of critical vowel for the multisyllabic English-language stimuli for U.S.-born speakers, as a function of place of articulation and [voice] value in the post-vocalic environment. .........................85

Figure 3.6 Mean duration (and standard error) of post-vocalic consonant closure for foreign- and U.S.-born speaker groups, as a function of language of utterance and [voice] value in the post-vocalic environment. .........................89

Figure 3.7 Mean duration (and standard error) of post-vocalic consonant closure for foreign- and U.S.-born speaker groups for the multisyllabic English stimuli and a subset of the Italian stimuli, as a function of language of utterance and [voice] value in the post-vocalic environment. .........................91
Chapter 1: Introduction

The objective of the research reported in this dissertation is to quantify the extent to which pre-consonantal vowels produced by bilingual speakers in each of their languages differ in duration to reflect language-specific settings triggered by the local segmental environment. Assuming that details of vowel duration fall under what House and Fairbanks (1953) termed “secondary acoustic characteristics,” i.e., those that are largely a matter of phonetic implementation, we ask whether it is possible for bilinguals to develop and maintain different timing settings for each of their languages, and whether (or under what circumstances) native-like settings are attainable in the second or less dominant language, in particular. That is, we use the timing variation present with vowel duration cross-linguistically as the medium through which to explore a phonetic aspect of bilingualism.

Numerous acoustic-phonetic studies have established that the environment of a vowel influences details of its acoustic realization. Of particular interest is the variation in vowel duration induced by voicing values in adjacent consonants. Studies of English have shown that the duration of a vowel preceding voiced consonants is greater than that preceding voiceless consonants (House & Fairbanks 1953, Denes 1955, Peterson & Lehiste 1960, Sharf 1962, Klatt 1973, Hillenbrand, Clark & Nearey 2001, Tauberer 2010, among others), and that this duration difference can even serve as a cue to the voicing quality of the following consonant (Denes 1955, Raphael 1972). We focus here on the influence of the post-vocalic consonant and the resulting consonantal voicing effect: A vowel is produced with greater physical duration, all else being equal, when the post-vocalic consonant is voiced as in *rib*, *wed*, or *bag*, rather than voiceless as in *rip*, *wet*, or *back*. This effect has consistently been reported to be markedly greater in English than in other languages (Zimmerman & Sapon 1958, Chen 1970, Mack 1982).
The lack of research on the production of vowels by bilinguals in their first language has been remarked upon by Flege (1995). Notably, many of the studies that examine vowel production by bilinguals in their second language focus on other aspects, such as spectral patterns, and utilize native-speaker judgments to assess the degree of “foreign accent” exhibited by non-native speakers (see Davidson 2011, which offers a concise summary of research on vowel production in the second language). The study reported here instead uses the phonetic setting for vowel duration, a timing indicator, as a measure of native-like ability. Using the consonantal voicing effect (CVE)\(^1\) as the vehicle for exploration, this project complements current research on the closely linked area of perception (e.g., Flege, Munro & MacKay 1995, Munro, Flege & MacKay 1996, Ingram & Park 1997, Piske, Flege, MacKay & Meador 2002, and others) and contributes to the knowledge base by providing data for Italian-English bilinguals using materials designed to sample the languages systematically while adding to our understanding of a potential source of foreign accent using vowel duration ratio as a marker.

The discussion which follows in this chapter first explores the CVE and the cues resulting from this vowel duration difference as well as briefly summarizes possible explanations for the effect. We then discuss bilingualism and second language acquisition, examining past studies that have looked at issues such as the Critical Period Hypothesis and the extent of second language learners’ fluency in their non-native language. We review models proposed by Flege (1995), Best (1995) and Best and Tyler (2007) which attempt to account for a number of phenomena in second language perception and production. These theories along with empirical evidence from prior studies on related topics combine to inform predictions on the question of how the CVE might manifest itself in each of a bilingual’s two languages. Considerations

---

\(^1\) Other terms used to refer to this same phenomenon include *vowel length variation* (e.g., Chen 1970), *vowel-length effect* (e.g., Kluender, Diehl and Wright 1988), *voicing-dependent vowel duration* (e.g., Hussein 1994) and *preceding vowel duration* (e.g., Tauberer 2010).
entering into the experimental implementation are discussed in Chapter 2, results of the study are reported in Chapter 3 and concluding remarks are presented in Chapter 4.

1.1 Vowel duration

Various studies have examined the correlation in English between consonants and their preceding vowels. Early investigations additionally explored the role of vowel duration as a perceptual cue to the voicing of the post-vocalic consonant. An important finding emerged for English: a lengthened vowel serves as a cue to the listener that a voiced, rather than voiceless, consonant follows, as in *cab*, cf. *cap*. Later work attempted to account for this durational difference between the voiced and voiceless environments with explanations couched in either articulatory or perceptual terms. This collection of studies led to another finding of interest: the durational difference between vowels preceding voiced consonants and vowels preceding voiceless consonants is markedly greater in English than in other languages. The following sections review a number of the studies that have examined the CVE, both for English and other languages, give a glimpse at the role of vowel duration in perception and look briefly at potential sources of the CVE.

1.1.1 Consonantal voicing effect in English

In their 1953 study, House and Fairbanks were interested in the impact that consonants varying in voicing, place and manner of articulation had on the vowel. They asked 10 speakers of American English to produce 72 disyllabic nonce words. The stressed vowel, which fell in the second syllable, was one of 6 vowels, */i,e,æ,a,o,u/*, and was surrounded by one of 12 consonants, */p,b,t,d,k,g,f,v,s,z,n,m/*, e.g., *hubeeb, hupeep*. Measurements were taken for vowel duration,
fundamental frequency and relative power. Effects of voicing, manner of articulation and place of articulation were all examined. Voicing was shown to have the greatest impact on the vowel, followed by manner of articulation then place of articulation. The voiced-voiceless difference for the set of oral stops was 83 milliseconds (ms), equal to a 1.52 ratio. Duration in the voiced context was always significantly greater than in the voiceless context.

Denes (1955) asked speakers of English to produce a word pair identical except for the voicing quality of the final segment – [juz] as in (to) use and [jus] as in (the) use. These words constitute a minimal pair, words that differ by just one phoneme, i.e., z versus s. Denes reported that tokens with the voiced consonant exhibited vowel durations ranging from 120 ms to 200 ms, while those with the voiceless consonant, 40 ms to 80 ms. Examining the smallest possible duration difference (120 ms for the voiced context, 80 ms for the voiceless), we may summarize the relative vowel duration across environments via the ratio 1.5. The voiced-to-voiceless ratio allows us to abstract away from speech rate variation and intrinsic differences in vowel length, providing a convenient means of comparison when materials and subjects differ.

A study of English by Peterson and Lehiste (1960) used a reading task as well, but this time the target words were uttered inside a carrier phrase, Say the word ____ again. Peterson and Lehiste had two data sets, one read by a single speaker and the second read by five speakers. Although not explicitly stated, all speakers are presumed to have been native speakers of American English. The first data set consisted of 1263 monosyllabic words with consonant-vowel-consonant structure (referred to as “CNC” by Peterson & Lehiste). The second set contained 70 words, 30 minimal CVC pairs and 5 minimal disyllabic pairs (although no examples of the disyllabic words used in the study were reported, nor did any of the results appear to reflect anything other than the monosyllabic words). Results for 118 minimal pairs
differing in final consonant voicing, e.g., *bead ~ beat, side ~ sight*, as found in the first set of data showed average vowel duration to be 297 ms before a voiced consonant and 197 ms before its voiceless counterpart, equal to a voiced-to-voiceless ratio of 1.51. The average duration values for minimal pairs in the second set of data which contained the same contrast in [voice] value for the post-vocalic consonant were 291 ms preceding voiced consonants and 193 ms preceding voiceless consonants, also corresponding to a ratio of 1.51. Peterson and Lehiste, however, did not differentiate here between sonorants and obstruents in post-vocalic position.

The large corpus of words utilized for the study allowed the authors to identify patterns amongst the stimuli in the data, including a four-way contrast for manner of articulation (stops vs. fricatives) and voicing (voiceless vs. voiced), e.g., *right ~ ride, rice ~ rise*. The corpus contained nine such sets, and Peterson and Lehiste reported average vowel duration values for each of the four possible pairings, as listed below in Table 1.1.

<table>
<thead>
<tr>
<th></th>
<th>Voiced</th>
<th>Voiceless</th>
</tr>
</thead>
<tbody>
<tr>
<td>stop</td>
<td>280</td>
<td>184</td>
</tr>
<tr>
<td>fricative</td>
<td>376</td>
<td>228</td>
</tr>
</tbody>
</table>

Values from just this subset correspond to a voiced-to-voiceless ratio of 1.52 for the stops and 1.65 for the fricatives. Peterson and Lehiste used additional data from their corpus, 21 monosyllabic word pairs in total, to further explore the effect of manner of articulation while voicing remained constant. The duration preceding voiced stops was measured at 300 ms (slightly higher than the value for the nine sets above) while the measurement preceding voiced fricatives was 379 ms (nearly identical to the value listed in the table above). Peterson and
Lehiste additionally examined minimal pairs, 68 pairs in total, that differed in [voice] value for the prevocalic consonant, concluding that the first consonant of the sequence played no discernible impact on vowel length\(^2\) – “the influence of an initial consonant on the duration of the syllable nucleus followed no simple regular pattern” (p. 701).

Sharf (1962) was interested in disyllabic words with a coronal consonant in intervocalic position, e.g., *ladder ~ latter*. In particular, he was interested in whether the coronal stops or their preceding vowels were instrumental in differentiating the word pairs, i.e., conveying the underlying [voice] value of the flapped segment. In order to have a basis of comparison, Sharf had a single speaker produce both monosyllabic and disyllabic real words, e.g., *cad ~ cat, caddy ~ catty*, which were usually (near-)minimal pairs contrasting for [voice] value of the post-vocalic consonant. For the disyllabic stimuli, stress was on the first syllable and the relevant consonant was in intervocalic position.

With the monosyllabic stimuli, Sharf reported a mean difference of 96 ms, equal to a voiced-to-voiceless ratio of 1.53. With the disyllabic stimuli, the difference was about 33 ms (1.28) for the non-coronal cases and 9 ms (1.06) for the coronal, i.e., flapping, cases. For the non-coronal disyllabic stimuli, Sharf found the closure duration to be greater when the [voice] value for the post-vocalic consonant was voiceless rather than voiced. He found the opposite trend with the coronal stimuli, where the closure was greater for the underlyingly voiced

\(^2\) But see discussion of Hillenbrand, Clark and Nearey (2001) later in this section for conflicting conclusions regarding the impact of the prevocalic consonant. Further, it should be noted that Peterson and Lehiste (1960) provided values both including and excluding aspiration from the vowel duration measurement when trying to determine the impact of the initial consonant. Values reported for the 68 minimal pairs showed the vowel to be on average 274 ms following a voiced consonant and 308 ms following a voiceless consonant when including aspiration (251 ms when excluding aspiration). Vowel durations were found to be greater following /s/, cf. /z/. The conclusions drawn by Hillenbrand et al. were based on oral stops; it would seem that Peterson and Lehiste could have come to the same conclusion if they limited their examination to oral stops and excluded the aspiration period from the measurement duration.
segment. He concluded that the cues, i.e., vowel and closure duration, were “reduced” for the coronal pairs in intervocalic position.

Similar ratios for monosyllabic and disyllabic stimuli were found by Klatt (1973). He had three speakers read words inside a carrier phrase, Say ____ instead. Stimuli consisted of 40 monosyllabic and 40 disyllabic words. A difference from the Sharf study is that Klatt included coronal place of articulation with his disyllabic stimuli. Klatt reported a ratio of 1.5 for monosyllabic words and 1.27 for disyllabic. Klatt was additionally interested in the magnitude of the vowel duration difference as a function of syllable count, and found that the [voice] value of the post-vocalic consonant corresponded with differing magnitudes: duration differences were greater when the post-vocalic consonant was voiced, e.g., room ~ rumor and lesser when it was voiceless, e.g., keep ~ keeper.

Hillenbrand, Clark, and Nearey (2001) sought to determine the effect of surrounding consonants on an intervening vowel in CVC syllables by asking subjects to identify the vowels present in these syllables when heard aloud. Although Hillenbrand et al. were mostly interested in changes to the spectral patterns and their impact on vowel identification, they also reported their findings for vowel durations for the CVC syllables. They recorded 12 speakers producing CVC syllables with every combination of /b,d,g,p,t,k/ in pre- and post-vocalic position, and /i,ɪ,ɛ,æ,a,ʌ,o,u/ in vowel position, e.g., [bib], [dup], and [tæg]. Hillenbrand et al. divided the utterances into four groups based upon [voice] value of the initial and final consonants. They found that, consistent with earlier findings, mean vowel duration was greater when the post-vocalic stop was voiced rather than voiceless. Contrary to the conclusions of Peterson and Lehiste (1960) mentioned previously, Hillenbrand et al. showed vowels to be longer, on the order of 20-40 ms, when the prevocalic consonant was voiced rather than voiceless, all else
equal. The table below lists the average durations reported for all vowels in each of the four groups.

**Table 1.2** Average vowel duration, in milliseconds, by consonantal environment, where \[-VD\] = voiceless stop and \[+VD\] = voiced stop, as reported by Hillenbrand et al. (2001).

<table>
<thead>
<tr>
<th>Environment</th>
<th>Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>([-VD] _ _ _[+VD])</td>
<td>146</td>
</tr>
<tr>
<td>([+VD] _ _ _[-VD])</td>
<td>177</td>
</tr>
<tr>
<td>([-VD] _ _ _[+VD])</td>
<td>224</td>
</tr>
<tr>
<td>([+VD] _ _ _ [+VD])</td>
<td>252</td>
</tr>
</tbody>
</table>

We calculate the overall voiced to voiceless ratio as 1.47 (76 ms), further divided as 1.53 for words with a prevocalic voiceless consonant and 1.42 for words with a prevocalic voiced consonant.

In a production study, Tauberer (2010) looked specifically at CVE magnitude in differing syllable structures, including environments in which flapping occurs. He asked five native speakers of English to read target words, (near-)minimal pairs contrasting for [voice] value in post-vocalic position, inside the carrier phrase, *Say ____ for me*. Target words were either monosyllabic, e.g., *thought*, or in one of two disyllabic conditions based on location of the trigger consonant, tautosyllabic, e.g., *seatbelt*, or heterosyllabic, e.g., *seater*. The stimuli included a range of vowels together with a post-vocalic consonant from the inventory of stops in English.

Tauberer found the voiced-to-voiceless ratios to be greatest for the monosyllabic stimuli, with a median value of 1.52, and smallest for the “flap target words,” with a median value of 1.09. For the disyllabic stimuli, excluding the potential flap cases, Tauberer reported a ratio of 1.27 (values taken from the left-hand panels of Figures 5.2 and 5.3 on p.51 of Tauberer [2010]).
He found no difference due to trigger consonant position (tautosyllabic vs. heterosyllabic) in the disyllabic stimuli, when excluding the flapping context. A summary of vowel duration ratios across studies for English follows in Table 1.3.

**Table 1.3**  Voiced to voiceless ratios for English, with corresponding study.³

<table>
<thead>
<tr>
<th>Ratio</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.47</td>
<td>Hillenbrand, Clark &amp; Nearey (2001)</td>
</tr>
<tr>
<td>1.50</td>
<td>Denes (1955)</td>
</tr>
<tr>
<td>1.50</td>
<td>Klatt (1973)</td>
</tr>
<tr>
<td>1.51</td>
<td>Peterson &amp; Lehiste (1960)</td>
</tr>
<tr>
<td>1.52</td>
<td>House &amp; Fairbanks (1953)</td>
</tr>
<tr>
<td>1.52</td>
<td>Tauberer (2010)</td>
</tr>
<tr>
<td>1.53</td>
<td>Sharf (1962)</td>
</tr>
<tr>
<td>1.57</td>
<td>Zimmerman &amp; Sapon (1958)</td>
</tr>
<tr>
<td>1.63</td>
<td>Chen (1970)</td>
</tr>
</tbody>
</table>

Ratios from the table above are graphed in Figure 1.1 below. The darker circles in the figure result from the overlap of identical ratios for separate studies.

**Figure 1.1**  Voiced to voiceless ratios for English, across studies.

1.1.2  **Consonantal voicing effect cross-linguistically**

Zimmerman and Sapon (1958) examined vowel duration in English and Spanish.

Spanish has a consonantal voicing contrast that distinguishes the three pairs \(p~b, \ t~d, \ k~g\); English also has these pairs (and a few others). Each participant — two native speakers of

³ Data from Sharf (1962), Klatt (1973) and Tauberer (2010) include the monosyllabic ratio only; see §2.1 for a discussion of default word structure. Ratios listed for studies by Zimmerman and Sapon (1958) and Chen (1970) are reported in the following section, §1.1.2.
American English and two native speakers of Spanish (from different areas of Latin America but with reportedly similar dialects) — read aloud a list of words in their native language, with each word produced in isolation. The English word list contained a total of 38 monosyllabic words selected to instantiate the contrasts in the final consonant and for presence of the vowel /i/, e.g., need, neat, seed, seat. The Spanish word list consisted of 90 disyllabic words with five different vowels in the stressed, penultimate syllable and with the post-vocalic consonant (located in the onset of the following syllable) contrasting for [voice] value, e.g., pido, pito, pago, pico.

Zimmerman and Sapon described the vowel duration effects they observed in the two languages as “qualitatively similar but quantitatively different” (p.152). That is, they found a greater magnitude for the CVE-English (83 ms, with average vowel duration of 228 in the voiced environment and 145 ms in the voiceless) than for the CVE-Spanish (18 ms, 127 in the voiced environment and 109 ms in the voiceless). The ratio for English was 1.57 while only 1.17 for Spanish.

We must consider a few items, including disparity in contexts, when evaluating this study. First, the target vowel and post-vocalic consonant are tautosyllabic in the English stimuli, i.e., they comprise the coda of a single syllable. In the Spanish words, however, the target vowel and post-vocalic consonant are heterosyllabic, i.e., the vowel fills the role of coda for the initial syllable while the consonant forms the onset of the following syllable. We know for English that, all else being equal, vowels are longer in open syllables than in closed syllables and they are longer in monosyllabic words than in disyllabic words (Ladefoged, 2006). We know for Spanish that voiced stops spirantize in intervocalic position. This is particularly noteworthy given that vowels in English tend to be longer preceding fricatives than stops (House & Fairbanks 1953,
Finally, this study used both obstruents and sonorants when measuring durations and calculating CVEs for the voiced and voiceless contexts.

As mentioned previously, many researchers were interested in a possible motivation for the consonantal voicing effect, including Chen (1970) who examined four languages: English, French, Korean and Russian. He first attempted to determine whether this variation was language-specific, or if in fact it appeared to be a language universal and would therefore, presumably, have an articulatory basis (see §1.1.4 for further discussion of this topic, including the mechanics of the effect as well as its status cross-linguistically). Chen used word lists composed of minimal or near-minimal pairs (7 pairs each for English and Korean, 10 for French, 11 for Russian). For English, three tokens per word were recorded by a single speaker. For French, Korean and Russian, a single subject for each language read each stimulus word aloud six times for recording, three times in isolation and three times together with the other member of the minimal pair.

Chen found that “in all four languages studied[,] a vowel is invariably longer before a voiced consonant than before an unvoiced one” (p.135). Additionally, the duration difference was significantly greater for English than for the three other languages, in accordance with Zimmerman and Sapon’s 1958 findings for English and Spanish. Chen observed the following voiced-to-voiceless ratios (and mean duration differences) for the four languages: English 1.63 (92 ms); French 1.15 (53 ms); Russian 1.22 (29 ms); and Korean 1.31 (28 ms). Again, we have a disparity in contexts. The English stimuli had one disyllabic word pair, *amble ~ ample*; the remaining pairs were monosyllabic. The French and Russian stimuli included both monosyllabic and disyllabic word pairs. The Korean stimuli contained only disyllabic words. Additionally,

---

4 Assuming that vowels tend to be longer preceding fricatives versus stops in Spanish as well, this would result in a greater ratio than expected if a stop were to surface. If, however, the segment is a stop underlyingly and the duration is not impacted by the surface form, then the change in manner of articulation would be irrelevant.
final consonants devoice in Russian and thus surface as phonetically voiceless. All of these factors must be considered when evaluating the results of the Chen study.

Table 1.4 below includes vowel duration measurements from the study for one minimal pair from each language, selected for comparison here based upon the relative similarity of the pairs.

Table 1.4  Vowel duration measurements, in milliseconds, as a function of [voice] value in the post-vocalic environment, for a single word pair in each of the four languages as reported by Chen (1970).

<table>
<thead>
<tr>
<th>Language</th>
<th>Voiced</th>
<th>Voiceless</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>/læg/</td>
<td>/læk/</td>
<td>145</td>
</tr>
<tr>
<td>French</td>
<td>/væg/</td>
<td>/læk/</td>
<td>22</td>
</tr>
<tr>
<td>Russian</td>
<td>/lug/</td>
<td>/luk/</td>
<td>36</td>
</tr>
<tr>
<td>Korean</td>
<td>/aga/</td>
<td>/aka/</td>
<td>28</td>
</tr>
</tbody>
</table>

Chen found vowel durations for French to be greater than for any of the other languages, but the CVE magnitude was much less than that found for English.

Esposito (2002) cited earlier studies of Italian (Vagges, et al. 1978, Magno-Caldognetto, et al. 1979) as having reported that Italian vowels have great duration before and after voiced, versus voiceless, consonants. She noted that both studies, however, lacked indication of whether their findings were significant. She then conducted her own experiment to test this: seven native speakers of Italian produced the carrier phrase, *Prendi ___ se vuoi*, ‘Take ___ if (you) want.’ The stimuli were VCVC nonsense words where the two vowels and the two stop consonants were identical (e.g. *abab, opop, utut*). Consonants came from the full inventory of oral stops in Italian, /b,d,g,p,t,k/, and vowels from the full set in Italian /i,e,a,ɔ,o,u/. Speakers produced three repetitions of 42 of these sentences.
Esposito found that the voicing feature of the consonants impacted duration of the stressed vowel. She did not specify or suggest whether the length effect came primarily from the preceding or from the following consonant but rather spoke specifically of voiced and voiceless “contexts.” Esposito’s data showed the average duration difference to be 28 ms (152 ms in the voiced context, 124 in the voiceless context), with a voiced-to-voiceless ratio of 1.23. Esposito’s findings confirmed the presence of the CVE in Italian, but with a much smaller duration difference than has been reported for English. A methodological concern as noted also by Esposito is that the stimuli produced in the study were inconsistent with the syllabic structure of Italian, a language where words typically end in a vowel, with few exceptions, e.g., loanwords, proper nouns.5 The finding by Hillenbrand et al. regarding the impact of the [voice] value of the prevocalic consonant on vowel duration as reported in the preceding section is relevant to the Esposito study, where the pre- and post-vocalic consonants were identical and thus matched for [voice] value. Since [voice] value of the prevocalic consonant was not held constant in the Esposito study, the potential impact of this consonant must be considered. A summary of vowel duration ratios for languages other than English where the CVE is known to be present follows in Table 1.5.6 Languages for which presence of the CVE is disputed will be discussed in §1.1.4.

<table>
<thead>
<tr>
<th>Language</th>
<th>Ratio</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>1.15</td>
<td>Chen (1970)</td>
</tr>
<tr>
<td>Spanish</td>
<td>1.17</td>
<td>Zimmerman &amp; Sapon (1958)</td>
</tr>
<tr>
<td>Russian</td>
<td>1.22</td>
<td>Chen (1970)</td>
</tr>
<tr>
<td>Italian</td>
<td>1.23</td>
<td>Esposito (2002)</td>
</tr>
<tr>
<td>Korean</td>
<td>1.31</td>
<td>Chen (1970)</td>
</tr>
</tbody>
</table>

5 The design of Esposito’s study appears similar to that of House and Fairbanks (1953). An effort to best match the stimuli for the House and Fairbanks study may explain Esposito’s use of non-prototypical Italian word forms.
6 See also Tauberer (2010) for a more extensive list of voiced-to-voiceless ratios by language and corresponding study.
Ratios from the table above, together with those for English as taken from Table 1.3, are graphed in Figure 1.2, below. Ratios for languages other than English all fall within the range between 1.1 and 1.4. Ratios for English all fall within the range between 1.4 and 1.7. As before, the two darkest circles in the figure amongst the English-language ratios represent the overlap of identical ratios from separate studies.

![Voiced to voiceless ratios](image)

**Figure 1.2** Voiced to voiceless ratios, across studies for English and other languages. The lighter markers represent values for languages other than English, while the darker markers represent values for English.

We use these values as a basis for comparison while acknowledging the discrepancy in contexts and inconsistency in procedures both within and across studies. The CVE results in vowel duration being roughly 50% greater in the voiced context for English while maximally half the size for other languages, suggesting a phenomenon that is of detectably greater magnitude in English than in other languages. What may appear at first glance to be an intrinsically phonetic effect has been phonologically enhanced in English, creating a greatly exaggerated effect. The magnitude of the effect in English surpasses what has been seen in other languages for both mono- and multi-syllabic words.

### 1.1.3 Duration as a cue

The vowel duration differences described in the preceding section play an important role in speech perception. These differences serve as cues in English, essentially indicating the voicing quality of the post-vocalic consonant to the listener. Denes (1955) had 33 subjects listen
to the same monosyllabic word pairs used in his production task, [juz] as in (to) use and [jus] as in (the) use, which differed solely in the voicing of the word final consonant. The stimuli were synthesized with identical spectral patterns but with durational differences in both the vowel and the final consonant. Denes used all combinations of four vowel durations (50 ms, 100 ms, 150 ms and 200 ms) together with five consonant durations (50 ms, 100 ms, 150 ms, 200 ms and 250 ms) for a total of 20 test items.\(^7\) To provide a frame of reference for the rate of speech, items were played following the synthesized phrase Is it an s or z? Listeners identified which word they heard by writing s for (the) use or z for (to) use.

Denes found that the duration of the vowel and the final consonant relative to one another played a role in perceived voicing of the consonant. As vowel duration increased, so did the percentage of z responses. Consonants with the shortest durations had the largest percentage of z responses. Conversely, as consonant duration increased, so did the percentage of s responses. Fewer s responses were seen with the longer vowel durations.

Raphael (1972) also used synthesized stimuli as a means to examine perception of final consonant voicing based on preceding vowel duration. To conduct his experiment, Raphael used steady-state vowel durations ranging from 150 ms to 350 ms (chosen using the range reported for the vowels in speech samples for the words spoken in isolation) with 50 ms transitions into and out of the vowels. He placed the vowels inside each member of a minimal pair matched for place and manner of articulation but contrasting for [voice] value, i.e., stops: bed ~ bet; fricatives: dues ~ deuce; and clusters: caused ~ cost, pigs ~ picks.

Twenty-five participants in total heard the stimuli over a loudspeaker in a laboratory, and then indicated for each stimulus which member of the minimal word pair they had

---

\(^7\) There were actually 80 items, as the set of 20 was constructed with four different intensities (see Denes 1955 for further description). Following Denes, we report here just on those items at 20 decibel vowel-consonant intensity ratio.
heard. Raphael found “with one exception /ʃ/ ~ /ʒ/ and regardless of the voicing cues used in their synthesis, all final consonants and clusters were perceived as voiceless when preceded by vowels of short duration, and as voiced when preceded by vowels of long duration” (p.1298). Raphael, however, did not provide a baseline of what to consider a short versus long vowel. An examination of the results provided by Raphael illustrated a pattern in which relatively shorter vowel durations corresponded to a perceived absence of voicing and relatively longer vowel durations matched up with a perceived presence of voicing.

1.1.4 Motivation for variation in vowel duration

Numerous explanations have been proposed to account for the variation in vowel duration that occurs within many, if not most, languages. These tend to relate to either the articulatory/production factors or the auditory/perceptual factors. We will briefly explore both possibilities (see Kluender, Diehl and Wright (1988) for a more complete summary of studies on this topic). We will not, however, discuss whether the difference in duration occurs due to vowel lengthening preceding a voiced consonant or to vowel shortening preceding a voiceless consonant as it is beyond the scope of this study.\(^8\)

The tendency to attribute the CVE to production-based sources, a common explanation in the literature, may derive from the near-universal presence of this effect. As mentioned in §1.1.2, Chen (1970) set out with the goal of determining the source of “vowel length variation.” Having found this effect with English, French, Russian and Korean and having no evidence to the contrary, Chen speculated that the CVE was likely physiological in nature and therefore a

\(^8\) For further discussion of this subject, see Chang (2002) who argues for vowel shortening. Additionally, Esposito (2002) considers the possibility of vowel shortening in Italian on the basis of vocal fold activity, i.e., the time required for the vocal folds to go from closed to open as opposed to remaining closed.
language universal. After review of several possible hypotheses, Chen sided with the rate at which the transition from vowel to consonant closure occurs as the articulatory source of this effect. According to this scenario, the pressure inside the mouth in production of a vowel-stop sequence is greater for a voiceless segment than for a voiced one and requires more effort. Therefore, the transition time from vowel to closure is faster with a voiceless segment while the vowel continues for a longer duration of time preceding a voiced segment.

To account for the large duration difference between voiced and voiceless contexts found with English but not with the other languages, Chen tentatively concluded that the extent of the CVE was language-specific. He suggested that there might be a perceptual explanation for the phenomenon in English, stating that the magnitude of the CVE “may be seen as a perceptual device serving as distinctive function in the phonological system of the English language” (p.142). This follows from his discussion that the interplay of vowel duration and closure duration — where voiced consonants have longer vowel duration and shorter closure duration while voiceless consonants have the inverse relationship — serves to enhance the perceived differences in [voice] value of the post-vocalic consonant.

In fact, we propose that the CVE as found in many languages represents a phonetic effect. This phonetic effect appears to be a difference that is typically large enough to be perceived and reproduced but is below the level of active awareness of speakers. The large magnitude of the CVE in English is beyond the size of a simply phonetic effect and is thus considered to have been phonologized into the language, i.e., the basic phonetic effect has been phonologically enhanced in English. This follows from Chen who stated that “the extent, however, to which an adjacent voiced or voiceless consonant affects its preceding vowel durationwise is determined by the language-specific phonological structure” (p. 139).
Kluender, Diehl and Wright (1988) proposed an auditory motivation for vowel length differences, noting that speakers use the CVE in conjunction with the post-vocalic consonantal duration to “enhance phonological distinctiveness” (p. 153). Kluender et al. argue that the presence of a long vowel preceding a voiced consonant increases the perceived shortness of this consonant while a short vowel preceding a voiceless consonant increases the perceived length of the closure duration. In effect, speakers enhance the pre-existing durational contrast, making the distinction between the sounds greater. Notably, the authors express the belief that CVEs in English have been exaggerated given that past studies tended to compare monosyllabic words from English with disyllabic words from other languages. We recognize the need to address these concerns with respect to cross-language comparisons in a materials set, maintaining keen attention to aspects such as syllable composition, language timing, sound inventories and licit word structure.

The prevalence of the CVE cross-linguistically suggested to Chen that the effect is a consequence of the basic phonetic mechanism involved in the production of the particular sequence of segments, i.e., vowel plus stop. The transition from vowel gesture to consonant gesture is less complicated when the consonant is voiced as compared with voiceless, all else being equal, as no adjustments need to be made with respect to the voicing mechanism, i.e., vocal fold vibration remains constant.

As mentioned earlier, however, not all languages have been found to exhibit the CVE. This observation appears to contradict our claim that the CVE is a phonetic effect. Tauberer (2010) provided a summary of numerous studies conducted for other languages, including those for which the CVE has not been found, namely Czech, Swedish, Arabic and Polish.9

---

9 See Table 2.2 on p.12 of Tauberer (2010) for the full list of studies.
Czech, Swedish and Arabic all have a phonemic vowel length contrast. As such, a sizable increase or decrease in vowel duration may impact word meaning, i.e., speakers of those languages must control more carefully for vowel length than would be necessary for a language like English, where vowel length is not contrastive. Are changes to vowel duration that typically result from the consonantal voicing effect prohibited in languages with phonemic vowel length differences? We discuss this question by looking at the occurrence of pre-boundary lengthening in languages that similarly have a phonemic vowel length contrast.

Pre-boundary lengthening is the process whereby articulation slows down at the end of an utterance, leading to greater duration in the final segments. It has been claimed that this process occurs in most languages. While the presence of pre-boundary lengthening in languages with contrastive vowel length has been called into question, Hakokari et al. (2007) countered that the lengthening occurs also with languages of this type and provided evidence from Finnish. “Final lengthening (FL), the tendency to slow down articulation at the ends of utterances, has been observed in almost any language in which the matter has been adequately investigated. The only exceptions mentioned in the literature are quantity languages such as Finnish, Hungarian, Estonian, and Japanese… Even in these the tendency has been confirmed later on” (Hakokari et al. 2007:1105).

Shepherd (2008) looked at Japanese, a language that also has contrastive vowel length, and showed that, when more than one process is at play, the proportional difference in effect size may differ from the expected. Specifically, he showed that the typical ratio of long vowel to short vowel for Japanese decreases when the vowels are subject to lengthening based on placement inside an utterance in pre-boundary position, as compared with a control set. In these
cases, the short vowels increased to a greater degree (in terms of percentage) than did the long
vowels, leading to a smaller long-to-short ratio.

Hockey and Fagyal (1998) found that pre-boundary lengthening occurs with Hungarian,
another language that employs phonemic vowel length. It seems likely that languages with pre-
boundary lengthening are more limited in the amount of vowel duration variation that may
appear, as large changes may impact meaning. Accordingly if the CVE occurs within languages
with phonemic vowel length contrasts, as we believe is the case, the CVE magnitude would
likely be restricted to prevent, for example, incorrectly recognizing a short vowel in front of a
voiced consonant as a long vowel.

Returning to the languages where the presence of the CVE has been rejected, we note that
a conflicting finding for Arabic was reported by Hussein (1994). Hussein provided evidence
showing that the CVE exists for Standard Arabic, albeit the effect was minimal. That is, even
with a language whose vowel duration is limited by contrastive vowel length, the CVE is
present. This finding supports the view of the CVE as a basic phonetic effect. It is worth noting
that the list of studies reported by Tauberer (2010) includes Dutch, a language with contrastive
vowel length that also exhibits the CVE. It may be that further study of Czech and Swedish
would similarly reveal the presence of the CVE, albeit with a similarly small magnitude as
reported for Arabic.

Turning now to Polish, the final language claimed to not exhibit the CVE, it is notable
that the language has word-final devoicing. This, however, may have had no impact on the
results founds. Other factors may have been at play that prevented the CVE from emerging or
being recognized.\textsuperscript{10} While the possibility exists that the CVE is not a consequence of a basic

\textsuperscript{10} It is also possible, as suggested by Hussein (1994) who addressed these issues in his own study of Arabic, that this
study lacked sufficient depth and breadth.
phonetic mechanism and is therefore not a language-universal, it seems that a more thorough examination of the languages in question — those for which it is claimed the CVE is not present — is necessary to shed further light on the cross-linguistic nature of the CVE. We now turn our attention away now from the particulars of the CVE and into a broader discussion of bilingual speakers.

1.2 Second language acquisition / bilingualism

We move now from a discussion of vowel duration into an exploration of second language acquisition and bilingualism with particular emphasis on the area of pronunciation. Examining phenomena such as foreign accent provides insight into the phonology or multiple phonologies of a bilingual. Whereas some researchers have strict definitions of what it means to be bilingual, perhaps only grouping balanced bilinguals (individuals who perform equally well in each of their languages) under the heading of bilingual, this way of thinking is less common nowadays. Similarly, we use a less constrained definition, encompassing all those who communicate effectively in two languages.

We follow the Grosjean (1989) perspective of a wholistic versus a monolingual view of bilingualism. The monolingual view considers the bilingual as composed of two monolinguals in a single person and compares the bilingual with actual monolinguals. The wholistic view instead assumes the bilingual to have a unique linguistic composition based upon these two languages, i.e., a configuration in which each language exists and functions in the presence of the other, in a complex linguistic architecture. Accordingly, Grosjean argues that the two languages of a bilingual do not get used interchangeably but rather get applied based on the situation or context of the speaker/hearer, contrary to the monolingual view.
Researchers tend to compare second language learners to monolinguals. We acknowledge the need to have a baseline of monolingual performance (as provided in the present study by previous studies), but also understand the benefit of comparing second language learners to bilinguals. Researchers working with bilinguals (or multilinguals) need also determine the dominant language of the individual while taking into account potentially complex patterns that may occur – such as a person English-dominant in one aspect of grammar, but not in others. The degree to which bilinguals perform as monolingual speakers of each of their languages depends on multiple items such as the task being performed and language dominance, the latter in turn being determined by numerous variables including age, which we turn to now.

1.2.1 Age effects in acquisition

One of the critical factors with regard to second language acquisition remains the age at which the second language is learned. Note that we do not differentiate here between acquisition and learning. Variation exists in the theories proposed by different researchers in terms of most and least favorable conditions for language learning, yet there tends to be a consensus that earlier is better, i.e., earlier exposure to a second language usually leads to better performance in that language.

The widely cited Johnson and Newport (1989) study maintains that individuals experience a gradual (linear) decline in performance based on age at which learning begins, starting at about age seven and continuing through the maturational period, after which point performance levels are reduced and accompanied by high variability in attainment levels. This point in time is commonly referred to as the critical period. The Johnson and Newport study looked at 46 native speakers of Korean or Chinese, labeling the 23 who had come to the U.S.
before age 15 as early arrivals and the other 23 who arrived after age 17 as late arrivals. All
participants had arrived in the U.S. between the ages of 3 and 39 and had been in the U.S.
between 3 and 26 years at the time of testing. They listened to 276 sentences and judged these
items as grammatical or ungrammatical (approximately half the stimuli were grammatical and
half ungrammatical). These sentences tested the participants’ knowledge of syntax and
morphology through such means as word order and past tense and plural markers (see Johnson
and Newport 1989 for a thorough discussion of materials used).

Johnson and Newport found that the performance by those who had arrived between the
ages of three and seven did not differ significantly from that of native speakers. Those who had
arrived between ages eight and ten had significantly poorer performance than did the native
speakers and the earlier arrival group. Each later arrival group had performance levels
significantly less than that of the one before, i.e. the 17-39 age group did worse on the
grammaticality judgment tasks than the 11-15 age group which in turn did worse than the 8-10
age group. Johnson and Newport also reported the age of language immersion to be a better
predictor of performance than the age at which formal language instruction begins (although the
two at times occur together), as the latter may take place in the native country which would result
in a very different learning environment than learning in the host country.

Birdsong (2002) surveyed results from numerous studies and, contrary to the findings of
Johnson and Newport, found no evidence to support an age-defined limit to successful language
learning. He claimed that age effects are not limited to a particular chronological age span but
rather correspond to a continually reducing success level in language attainment. Birdsong
highlighted that, although not common, native-like proficiency is achievable among late learners.
Hakuta, Bialystok and Wiley (2003) used data taken from the 1990 U.S. Census for a very large number of respondents, all of whom were native speakers of either Spanish or Chinese, and found that age and education were each important predictors of success in the second language, English. Also in contrast to the findings of Johnson and Newport, Hakuta et al. found no evidence that performance differs significantly based on whether learning, which corresponded in this study to age of immigration, takes place before or after the putative critical period. We turn now to the specific area of pronunciation in second language learners.

1.2.2 Pronunciation as a function of age

A number of age-related studies focus on accent/pronunciation. Whereas proficient second language learners often master the syntax and the semantics of their non-native language, pronunciation can remain a challenge. A study by Oyama (1982a) used 60 Italian-born male immigrants to examine issues of age of arrival/acquisition (AOA) and length of residence. Age of arrival in the U.S. for these participants ranged from 6 to 20 years, while number of years in the U.S. stretched from 5 to 18 years. Participants read a short paragraph aloud in English and also provided a spontaneous speech sample. Both categories of speech samples were rated for degree of accent by native speakers of American English. In addition to finding a strong age of arrival effect, Oyama’s results also showed that participants were perceived as having less accent with non-scripted speech than with scripted speech.

Oyama (1982b) conducted a “masked” speech test where subjects (presumably the same 60 speakers in Oyama 1982a, although not explicitly stated) identified words produced by a native speaker of English after machine-created white noise had been added to the utterances. Participants scored points for their ability to recover words in these sentences. Ten native
speakers of English served as a baseline for performance. Those who started learning the language at or prior to 11 years of age performed like native speakers on this task. Oyama found that those who started learning English in early adolescence, ages 11 to 15, had lower scores than the group that began learning at a younger age; those who started learning later still, ages 16 to 20, earned much lower scores. Based upon this finding, Oyama argued that the terminology sensitive period better describes the gradual decline in learning that occurs than does critical period. She concluded that age of learning is a major predictor in degree of foreign accent, but length of stay is not.

Bongaerts, van Summeren, Planken and Schils (1997) compared three groups of speakers producing English utterances: native speakers of Dutch with high English language proficiency, all of whom began language instruction in English about or after the age of 12; native speakers of Dutch with varied English language abilities; and native speakers of British English, who served as a control group. A separate group of native speakers of British English served as judges and rated the spoken utterances for degree of accent. The stimuli contained some speech sounds common to both English and Dutch, and other sounds present in English but absent from Dutch. Although the Dutch speakers with high English language proficiency did not perform as well as the native speakers of British English, they performed well enough for Bongaerts et al. to conclude that late learners can develop native-like pronunciation in a second language. Factors they noted as contributing to more native-like performance included a high level of motivation as well as continual and frequent exposure to English.

Moyer (1999) attempted to tease apart some of the factors that may lead to highly successful performance in a second language, placing particular emphasis on the notion that age alone may not be a significant predictor and that age gets confounded with other aspects, such as
motivation and attitudes toward learning and success. This study examined 24 graduate students of German at the University of Texas, Austin, all of whom had numerous years of German language instruction and had spent time in Germany. The mean age of first exposure to German was 15.2 years of age. Moyer assumed all subjects were highly motivated based upon their course of study and their stated goals, as determined through a detailed questionnaire completed by participants. Four native speakers of German served as the control group.

Four judges, also native speakers of German, rated the participants on their production of three reading tasks (a list of 24 words, a list of 8 sentences and a paragraph of text) and a free-speak (spontaneous speech sample). Judges rated the speech samples as native or non-native, and provided a confidence level for each judgment on a three point scale. Results illustrated that non-native speakers were more likely to be judged as non-native speakers than native speakers. Later age of immersion tied in with more judgments of non-nativeness, i.e., sounding less native. Longer immersion in the target country, however, did not necessarily equate with more judgments of native-like speech samples. Non-native speakers exhibited little variation in their performance on each task. Once other factors such as attitude and motivation were separated out, age was not found to have the great impact that other researchers had suggested. Finally, participants who had received some level of supra-segmental training performed in a more native-like fashion than the others.

Munro, Flege and MacKay (1996) examined the production of Canadian English vowels by 240 native speakers of Italian who were long term residents of Canada, to see the relationship between age of second language learning and perceived accentedness, and whether some vowels are easier to learn than others. The native speakers of Italian were aged 2 to 23 at the time of
immigration to Canada and had spent an average of 32 years in Canada at the time of testing. The control group consisted of 24 native speakers of Canadian English.

Participants heard a target $bVC$ word, e.g., *bit, beat, book*, inside the phrase, ___ is the next word, and then produced the word in the carrier sentence *Now I say ___*. Ten native speakers of Canadian English judged the vowels for degree of foreign accent. At the time of listening, judges were told the word to use as their basis for comparison in determining amount of accent, e.g., *bit*. To minimize the possibility of judges perceiving an Italian accent based on the surrounding consonants, the final consonant was removed from each word and the initial consonant was manipulated, due to presence of prevoicing in Italian [b] as compared with English [b]. Native speakers rated the production of the late-arrivers to be more accented than that of native English speakers, with the degree of perceived accent increasing as a function of increased age of learning. No specific cut-off age was indicated to separate the participants into early and late arrival groups. Some vowels seemed easier to learn, but no single vowel was perceived to have been produced in native-like fashion by all subjects.

In a second experiment, a vowel identification task, native English speakers named the vowel produced by the native speakers of Italian. Non-native speakers obtained much higher scores for this task as compared with the foreign accent rating task, illustrating that speech may be accented, yet still highly intelligible.

1.3 Models of second language speech perception and production

To better understand the factors involved in language usage by native versus non-native speakers, various models of second language speech perception and production have been proposed. These theories aim to account for variation in individual pronunciation based upon
such issues as age of learning and differences in the phonetic categories of the first and second
languages (L1 and L2, respectively). One such model is Flege's Speech Learning Model
(SLM). According to Flege, adults possess pre-existing categories for vowels and consonants in
the native language inventory. When listeners receive auditory input from a new language, they
either assimilate the sounds to an existing category or dissimilate them from an existing
category; the latter may lead to establishment of a new category for the sound. Numerous factors
play into the option chosen, such as age of the learner and native and second language
inventories.

According to Flege (1995), the speaker/hearer maintains phonetic categories comprised
of language-specific speech sounds. When a listener receives new input that sounds very similar
to an existing phonetic category, in all likelihood that sound will be assimilated to the category.
A listener who hears a distinct difference between the L2 sound and the closest L1 sound can
create a new category for the sound. Obviously, the greater the difference between the two
sounds, the more likely the difference will be perceived and the phonetic differences
discriminated. Difficulty arises for the listener when two sounds in free variation in the native
language are used contrastively in the second language, e.g., Japanese /ɾ/ ~ /l/ distinction. Flege
also highlights the possibility for dissimilation of two relatively similar sounds. If the bilingual
attempts to maintain a phonetic contrast between two sounds which share a phonological space,
the L2 category may be pushed away from the L1 category to impose the distinction. A main
claim of Flege’s SLM states that phonetic categories do not freeze but rather can be altered
throughout one’s life for both L1 and L2 categories.

Best’s Perceptual Assimilation Model (PAM) (1995) addresses how listeners deal with
non-native segments as well as non-native contrasts. In this model, speech sounds are
understood to be the result of specific combinations (constellations) of articulatory gestures. According to PAM, a listener encountering a new speech sound may proceed in one of the following directions: assimilate the sound to a native category, whether as a good or bad example of the category; regard the sound as speech-like but as incompatible with the native categories; or deem the sound not to be a speech sound.

Best’s PAM also divides perception of non-native speech contrasts into the following six different patterns:

- Two-Category Assimilation – the two speech sounds assimilate to different native categories;
- Category-Goodness Difference – the two speech sounds assimilate to a single category, but one is a distinctly better example of the sound than the other;
- Single-Category Assimilation – the two speech sounds assimilate to a single category, but neither is much better/worse of an example of the sound than the other;
- Both Uncategorizable – while recognized as speech, neither sound falls within a native category (although one or both may fall close to a native category);
- Uncategorized versus Categorized – one sound assimilates to a native category, the other (while still recognized as a speech sound) does not; and
- Nonassimilable – neither sound gets categorized as a speech sound.

According to Best, adult listeners more easily discriminate sounds which fall into two different categories or those that are clearly distinguished within a single category than most of the other types. Best’s attention to the wide variation in discrimination of non-native speech contrasts based on sound type distinguishes her model from Flege’s SLM, with its ability to account for a greater amount of variation in data.

Best, McRoberts and Goodell (2001) build on this distinction by providing empirical support. They report an experiment in which native speakers of American English attempted to

---

11 See Browman and Goldstein (1992, 1995) for extensive discussion of this theory, Articulatory Phonology, as well as Hall (2010) for a recent review.
discriminate three pairs of Zulu contrasts. As predicted by the model, participants’
discrimination varied for the three pairs, where performance on the Two-Category Assimilation
pair exceeded that for the Category-Goodness Difference pair which in turn surpassed that for
the Single-Category Assimilation pair. In this study, Best et al. categorized the Zulu contrasts
into pattern types on the basis of “articulatory similarities and differences between the Zulu
consonants and the most closely corresponding AE consonants” (p.779). While PAM lacks an
independent measure for predicting category types for the non-native speech contrasts, SLM, as
pointed out by Flege himself (Flege 1995), lacks “an objective means for gauging degree of
perceived cross-language phonetic distance” (p.264).

PAM-L2 (Best & Tyler 2007) furthers the agenda of PAM, looking at “equivalence” in
both phonetic and phonological terms. Under this framework, Best and Tyler specify that two
sounds, one from the L1 and one from the L2, can be classified as equivalent on the phonological
level but perceived as different on the phonetic level. We understand this to imply for the
present study that a post-vocalic consonant that triggers one timing setting for vowel duration in
the L2 and a different timing setting in the L1 need not be considered different segments at the
phonological level.

1.4 Bilingual production

As indicated earlier, we wish to use vowel duration, a straightforward, language-specific
setting, as the means to explore the languages of the bilingual. We are interested in the phonetic
variation found within and between groups of bilingual speakers, remembering that large
differences may exist from one group of bilinguials to another with respect to language history
and usage. While a number of studies examine the issue of vowel duration in bilingual speakers,
many focus on perception or look solely at English language production by second language learners. We instead explore vowel duration production by bilingual speakers in both of their languages. We look here at a few production studies conducted with bilingual speakers, starting with cross-linguistic studies of inter-speech posture and voice onset time before moving onto vowel duration. Studies of language-specific settings which use different phonetic parameters allow us to look at the same general question and make predictions about the consonantal voicing effect in bilingual speech production.

1.4.1 Inter-speech posture

Wilson and Gick (2014) explored “language-specific articulatory settings” in bilingual speakers of French and English. They focused on the default position of the articulators, i.e., the inter-speech posture (ISP). ISP refers to the position of the articulators as occurs between gestures, a default setting which differs from rest position. In previous work, Wilson (2006) concluded that monolingual speakers of French and English have language-specific articulatory settings, with significantly different ISPs found for each language. Wilson and Gick (2014) followed up on that study with two stated aims. The first was to see if these language-specific articulatory settings were necessary for bilinguals to sound native-like when speaking in monolingual mode, i.e., when the speaker was aware of the language being employed throughout the task, whether French or English. The second goal was to determine the setting for a speaker when in bilingual mode, i.e., when the task employed French and English interchangeably, and to see if that bilingual mode setting was consistent with the setting for the speaker when in monolingual mode for one language or the other.
Wilson and Gick measured articulatory settings for eight bilingual speakers of Québécois French and Canadian English during production of sentences uttered in French or in English using ultrasound imaging and optical tracking. The mean age of the speakers was 30 years; they had usually been exposed to French and English at a fairly young age. All had varying degrees of exposure to the two languages at the time of the study (e.g., one participant reported having 10% exposure to French and 90% exposure to English in a typical week, while another reported 60% French and 40% English). Stimuli consisted of 180 sentences, broken into 6 blocks of 30 sentences each, which were presented on a screen, one at a time. Two blocks consisted of solely French language utterances (to achieve monolingual French mode) and two of solely English language utterances (to achieve monolingual English mode). The remaining two blocks contained a combination of French and English such that the speaker did not know which language to expect next on the screen (to achieve bilingual mode). Ultrasound imaging captured tongue movement, while optical tracking captured lip, jaw and head movement.

Wilson and Gick divided the eight French-English bilinguals into two groups of four based on how native-like each sounded, as perceived by native listeners. In one group, all speakers were perceived as native-like in each of the two languages. In the second group, all speakers were perceived as not native-like in at least one language (one was not perceived as native-like in either language, two sounded native-like in French only and one sounded native-like in English only). Wilson and Gick analyzed the measurements for the French and the English for tongue tip height and lower lip protrusion only, which they believed to be the most informative and most closely matched to native-like performance. Measures were analyzed for 419 rest positions, averaging approximately 52 per speaker.
Consistent with their hypothesis, Wilson and Gick found that bilingual speakers in the first group, i.e., native-like in both languages, had distinct, language-specific ISPs, corresponding to monolingual behavior in each language. All four of these speakers differed significantly across the two languages with respect to the measurements for lower lip protrusion, and two differed significantly for tongue tip height. These four speakers each had a separate ISP for each language. The second group, however, did not have the same distinction in articulatory settings for the two languages.

To interpret the results when a speaker was in bilingual mode, Wilson and Gick looked at just the participants deemed to be native-like in both French and English. The settings for the articulators while in bilingual mode always had a setting in common with at least one of the monolingual-mode settings. These speakers tended to have an ISP that matched up with the ISP for the language most commonly used by the speakers in their present-day lives.

The results of this study are particularly relevant. ISP settings serve here as a good proxy for native-like ability. All of the speakers perceived as native-like in both of their languages had a distinct phonetic setting for the ISP in each of their languages. Vowel duration is another type of phonetic setting that can be measured, in this case via the voiced-to-voiceless ratio. In much the same way, we would expect bilingual speakers of Italian and English to have a distinct phonetic setting for the vowel duration ratio in each of their languages if they are to be perceived as native-like. In the case of vowel duration, the ratio is achieved by the resulting difference in the setting as a function of [voice] value. In any event, since the setting for vowel duration differs across monolingual speakers of Italian and of English, we would expect to find some difference exhibited by bilingual speakers in each of their two languages. In their discussion,
Wilson and Gick acknowledge that articulatory settings may be more rigid than other phonetic settings and thus less likely to conform to the setting of a second language.

### 1.4.2 Voice onset time

Zampini and Green (2001) examined the role of voice onset time (VOT) in two languages, Spanish and English, as reported in Magloire and Green (1999). Spanish and English both have voiced stops /b,d,g/ and voiceless stops /p,t,k/. The two languages differ in their VOT values for voiced and voiceless stops, as shown in Table 1.6.

<table>
<thead>
<tr>
<th>Voiced</th>
<th>Voiceless</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish Prevoiced</td>
<td>VOTs &lt; 35 ms (short lag)</td>
</tr>
<tr>
<td>English VOTs &lt; 35 ms (short lag)</td>
<td>VOTs &gt; 30 – 35 ms (long lag)</td>
</tr>
</tbody>
</table>

VOT values for voiced stops in English overlap with those of voiceless stops in Spanish.

A pair of stops contrasting for [voice] value, i.e., /p/ ~ /b/, were produced by early Spanish-English bilinguals as well as by monolinguals in each of the two languages and VOT values were measured. Each of the bilingual speakers, all of whom used both languages on a daily basis, had Spanish as a first language and had begun learning English by age five.

Zampini and Green outlined three possible outcomes. First, the bilinguals could have VOT values that correspond with monolingual-like behavior in each of their languages. Given the differing VOT values in the two languages, this would indicate two separate representations on the phonetic level for each stop, one for each language. This corresponds to what Fernández
(2003) refers to as language dependence. As a second possibility, the bilinguals could have a merged setting for the VOT value, falling somewhere in between the values that monolinguals for each of the languages exhibit, utilized for production in both languages. This would seem to indicate one phonetic category. Finally, the bilinguals could have monolingual-like values in just one of the languages. Although Zampini and Green do not go into detail on this third outcome, this could result in either one or two representations, depending upon if production in the two languages was equivalent, i.e., one representation, or different, i.e., two representations. If the strategy used for both languages is the same, this corresponds to what Fernández terms language independence.

With respect to VOT measurements, Zampini and Green reported no significant difference in production by the monolinguals and the bilinguals in each of the languages, i.e., bilinguals maintained two separate phonetic representations for VOT. Zampini and Green also discussed duration of voiceless closure intervals for word initial /p/ and /b/ in the two languages, as reported in Green, Zampini and Magloire (1997). They found that bilinguals exhibited parallel, but not identical, behavior to monolinguals.12 This second finding is important because it shows, as pointed out by Zampini and Green, that the early bilinguals had monolingual-like behavior for one phonetic setting (VOT) but not for the other (closure interval). The existence of differing outcomes underscores the importance of viewing bilinguals as individuals with a unique composition, as follows from the Grosjean perspective.

Flege, Munro and MacKay (1995) looked at voice onset time for 240 native speakers of Italian living in an English-speaking part of Canada.13 Speakers had immigrated to Canada between the ages of 2 and 23 (with a mean of 13 years) and had spent between 14 and 44 years

---

12 See Zampini and Green (2001) for further detail regarding the portion of the study discussing closure interval.
13 This is the same group of speakers from Munro, Flege and MacKay (1996) as described earlier in the chapter, in §1.2.2.
(with a mean of 32 years) in Canada. Speakers were divided into groups based on the age at which they had begun learning English. Similar to Spanish, Italian speakers produce /p/ and /t/ with short lag VOT values. Italian /k/, however, is produced with a long lag value, as is the case in English with all voiceless stops. Participants were given a list of words in English. They heard the phrase, ____ is the next word, over a loudspeaker and repeated the word back in the carrier phrase, Now I say ____. Looking at voice onset time for the word-initial segments of [p]ick, [p]eak, [t]ack, [t]ag, [k]ap and [k]ab, Flege et al. found VOT values to be longer for /k/ than /t/, and longer for /t/ than /p/. The data showed that, as age of arrival in Canada increased for this group of speakers, mean VOT values decreased for /t/ and /p/, making them less similar to VOT values for native speakers of English and indicating a carryover effect for VOT from the Italian into the English. Flege et al. claimed that early learners — at least most of them — established new phonetic categories for /p/ and /t/ (/k/ would be interpreted differently given that native speakers of Italian already have a long lag value for this segment) but only some late learners did the same.

The table below compares three sets of VOT values from Flege et al. (1995): for Italian, as spoken by native speakers of Italian (as reported by Vagges et al. 1978); for English, as spoken by a control group of 24 native speakers of Canadian English; and for English, as spoken by the group of Italian-English speakers whose mean AOA was 21 years of age.

<table>
<thead>
<tr>
<th>Language of utterance</th>
<th>/p/</th>
<th>/t/</th>
<th>/k/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native Italian</td>
<td>12 ms</td>
<td>17 ms</td>
<td>50 ms</td>
</tr>
<tr>
<td>Native Canadian-English</td>
<td>57 ms</td>
<td>78 ms</td>
<td>77 ms</td>
</tr>
<tr>
<td>Italian-English bilinguals</td>
<td>35 ms</td>
<td>45 ms</td>
<td>65 ms</td>
</tr>
</tbody>
</table>
The late learners of English had VOT values which appeared to be a merged setting between the values of the other two groups. We lack the corresponding VOT values for Italian for these bilingual speakers, information which would tell us if a compromised setting was used for both languages, or just for one.

1.4.3 Vowel duration in the second language

As part of the same study that looked at VOT, Flege, Munro and MacKay (1995) provided vowel duration values for Italian-English bilinguals as well as a control group of native speakers of Canadian English from their production of two minimal pairs in English. In the first pair, the trigger consonant was tautosyllabic, i.e., tag [tæg] ~ tack [tæk]; in the second pair, it was heterosyllabic, i.e., tagging [tæ.gɪŋ] ~ tacking [tæ.kɪŋ]. The morphological alternation was identical in each of the pairs.

Flege et al. found mean vowel duration to be greater in monosyllabic versus disyllabic words and greater preceding voiced versus voiceless stops (197 ms for /g/, cf. 157 ms for /k/), equal to a voiced to voiceless ratio of 1.25. See Appendix A for the corresponding figure from Flege et al. Additionally, the ratio appeared to be greater for monosyllabic words than for multisyllabic words. Although Flege et al. had a large number of speakers, the stimuli consisted of a set of just two minimal pairs and thus we must proceed with caution when making predictions on the basis of these results.

It is worth briefly discussing closure duration, values for which were also included for the two word pairs by Flege et al. Examination of the closure data (also shown in Appendix A)

---

14 Mean durations were reported as 206 ms for one syllable words and 248 ms for two syllable words. Looking at the accompanying figure, as shown in Appendix A, it appears that the duration value for the two syllable words might instead be 148 ms.
revealed that native speakers of Italian (at all AOAs) exhibited the same general pattern as found with native speakers of English: closure durations were greater preceding voiceless as compared with voiced consonants. Speakers with increased AOAs (11 and older) had significantly greater closure durations than was found with the native speakers of English.

Bent, Bradlow and Smith (2008) looked at vowel production by 10 native speakers of Mandarin. The speakers were university students, newly arrived in the U.S. after a period of formal study of English in their home country, which averaged 10 years. Vowel duration production was examined in two contexts, both of which are absent from the native language. The first was the tense/lax distinction: Mandarin is considered to have only tense vowels. The second was the difference in duration based on [voice] value of the post-vocalic consonant. This distinction does not exist in Mandarin, where syllables are either open, or closed by a nasal. A control group of 10 native speakers of English, all university students, also produced the stimuli.

No significant difference was found between the two groups with the tense/lax duration contrast. However, native speakers exhibited a much greater CVE than non-native speakers; the voiced-to-voiceless ratio was 1.54 for the native speakers of American English and 1.17 for the native speakers of Mandarin. The authors concluded that in this instance the CVE setting had not transferred over from the L1 but was learned by the Chinese speakers. The authors also concluded that certain temporal patterns are easier for L2 speakers to learn and produce than others. They considered the tense/lax distinction an easier pattern to learn than the CVE.

1.4.4 The present study

In this investigation, we ask whether native speakers of Italian who were late-learners of English have maintained the language-specific phonetic settings for the CVE found with Italian
monolinguals, and whether they have transferred these settings into their production of English words, consistent with the findings of Flege et al. (1995) for VOT production by late-learners. We also ask whether early learners of Italian and English have acquired two different timing settings for the CVE, corresponding to monolinguals in each of their languages. To address these questions, we designed a carrier phrase production task to assess the CVE as produced by our two groups of bilingual speakers in both Italian and English.

By way of background, we provide some basic information about Italian as it relates to English, with particular attention to the task at hand. The descriptions laid out here for Italian draw largely from Krämer (2009) and have been cross-checked against other sources. Italian and English overlap in their inventory of stops: /b,d,g,p,t,k/. A substantial difference between the two languages is that the coronal stops are subject to flapping in intervocalic position in English, but not in Italian.

Italian has seven vowels in stressed syllables, /i,e,ɛ,a,ɔ,o,u/. This entire inventory can be found within the considerably larger set of English vowels, with a primary difference being that /e/ and /o/ are diphthongized in English. Italian has a five-vowel inventory in unstressed syllables as the tense/lax distinction for mid vowels, /e,ɛ/ and /o,ɔ/, is neutralized in this environment: unstressed mid vowels must be tense.

Neither Italian nor English has a phonemic vowel length distinction. Given that length is not contrastive in Italian vowels, some cross-linguistically common generalizations about vowel duration in the two languages can be made. First, vowels in open syllables have greater duration than those in closed syllables. Second, longer words — those that have more syllables — tend to have shorter vowels. In §1.1.1, §1.1.2 and §1.4.3 we reported studies illustrating that the CVE
was present for both Italian and English, and that the size of the effect differed in the two languages.

Stress may fall on any syllable in Italian, but manifests itself on the penultimate syllable for the vast majority of lexical items. In the absence of a lexical stress mark, penultimate stress is the default for verbs. Nouns, however, lack a default stress setting. It may also be the case that penultimate stress assignment has emerged as the default placement given that the most frequently used words in Italian are disyllabic (Krämer 2009:198). Italian and English are considered to differ in their rhythmic timing patterns, where Italian is typically classified as syllable-timed and English as stress-timed (Dauer 1983, among others).

While the current study has Italian as the focus of study, and not an Italian dialect, the topic is worth mentioning. Dialect is pervasive in Italy and at times constitutes a separate language entirely. As Krämer writes, “Italians make a clear-cut distinction between *dialetto* and *italiano* and most speakers can be said to be bilingual in the sense that they have some competence in both a dialect and Italian,” adding that “the majority of speakers, even those who are not bilingual and don’t have any active command of a dialect, use a regionally ‘coloured’ version of Italian” (p.2). Consequently, despite the fact that two speakers of Italian can read the very same orthographic word (written standard Italian has a clear, established form), the pronunciation of that word may take disparate forms. These dialectal influences underscore the importance of attempting to recruit from a single region, to the extent possible, to avoid differences that might surface. Participant recruitment will be further addressed in §2.1 and, in particular, §2.2.1.

Considering the results across the bilingual-speaker studies discussed in this section, we expected our foreign-born bilinguals with late-acquired English to have a monolingual-like
setting for vowel duration in Italian and a merged setting for vowel duration in English.

Moreover, we expected our U.S.-born bilinguals with early exposure to both Italian and English to have a monolingual-like setting for vowel duration in English. Studies of vowel duration have left unanswered the question of how vowel duration would manifest itself in Italian for the U.S.-born speakers, a matter which we address in the present study.
Chapter 2: Implementation of the experiment

The project, as sketched in the preceding chapter, asks about the extent to which a bilingual speaker might acquire native-like command of phonetic settings for each of his or her languages. To answer this question, the plan is to compare the performance of Italian native speakers for whom English is acquired late and that of speakers who had early exposure to both Italian and English. The study assesses the duration of pre-consonantal vowels in elicited utterances as produced by the two groups of Italian-English bilinguals, based on a well-established cross-linguistic variation in the magnitude of the consonantal voicing effect (CVE) discussed in §1.1. The idea is to use that effect’s magnitude in each of a bilingual speaker’s languages as a proxy for native-like ability.

Although considerable research effort has been expended investigating the CVE across languages, there is a surprisingly scant literature, in particular, on the cross-linguistic performance of bilingual speakers. In what follows, we first discuss the rationale behind the decisions that shaped the overall design of the project (Section 2.1), before moving on to specify the detail of the project’s execution with respect to participant selection, materials design, data management, and such (Section 2.2).

2.1 Project design considerations

We want to understand how contextually determined vowel durations are controlled, all else being equal, in response to the circumstances of language acquisition. Durational differences, as triggered by the voicing quality of the post-vocalic consonant, can be quantified via a ratio, i.e., a phonetic setting. CVE-effects manifest themselves with a distinct ratio for each
language, based upon the way in which vowel duration is manipulated in the given language. In one case, speakers would be deeply rooted in their first language (Italian) at the point of acquisition of their second language (English); with bilinguals of this type, we ask what phonetic settings are achievable in a second language when values have already been established for a first language, and whether the CVE-setting in the second language can differ from that in the first language. With bilinguals who experienced early exposure to two languages, it can be assumed that speakers are not tethered to an allegiance to one particular language. Here we can ask whether distinct, native-like phonetic settings are achievable in each language, or if there is instead a single compromise setting. When two languages are acquired simultaneously, is it possible to maintain two separate, language-specific settings for the CVE?

We refer to participants who are likely to have established duration settings for their native Italian that correspond to those of monolingual Italian speakers as foreign-born bilinguals, i.e., native speakers of Italian with late-acquired English. Such participants are to be contrasted with what we refer to as U.S.-born bilinguals, i.e., speakers who experienced simultaneous (or near-simultaneous) acquisition of Italian and English. At the same time, marked variation among Italian dialects had to be taken into account, because such variation in all likelihood impacts pronunciation. In the ideal, dialectal uniformity is sought, to the extent possible. The several constraints of these design demands were resolved by drawing speakers from a culturally active community in New York City, whose members were bound by personal or family affiliation to a single Italian geographical area.\footnote{The study could alternatively have been implemented in Italy, contrasting U.S.-born native speakers of English who had migrated many years earlier to Italy with Italian-born speakers who had experienced simultaneous acquisition of Italian and English. For logistical reasons, the research was carried out locally.} Participants in the experiment, regardless of bilingual type, had origins in Quaglietta, Italy (region of Campania). See Figure 2.1.
We recruited two distinct speaker subgroups from this community, foreign-born and U.S.-born. Foreign-born bilinguals had immigrated to the United States as teenagers or young adults, resulting in a considerable gap between acquisition of their first and second languages. While well-integrated into their new and English-majority society over a period of at least 30 years of immersion, their commitment to Italian language and culture had been strongly maintained within a close-knit Italian-American community, as well as within the home. The U.S.-born adult children of such immigrants provided a reasonably close approximation to speakers with two first languages. U.S.-born bilinguals varied amongst themselves with regard to the language used in wider social circles, in employment situations, and within the family, but

---

were uniform with respect to their language of education, which was always English. Participant
details are spelled out in §2.2.1.

A group-based design for the experiment was selected, given the greater distance
between the two groups than within either group in terms of language experience. Accordingly,
language history data became essential, providing the background information required to group
participants on the basis of the circumstances of their language acquisition and usage. Moreover,
language histories are broad-brush instruments that provide an avenue for informal explorations
of the data, examining variables beyond the scope of those included in the group-based design,
e.g., social contexts in which one or other language was more likely to be used, that might
potentially contribute to performance. Administration of a questionnaire allowed collection of
language histories for each participant. The questionnaire, similar to the Language Experience
and Proficiency Questionnaire (LEAP-Q) (Marian, Blumenfeld & Kaushanskaya 2007) and other
such instruments, was adapted from that of Fernández (2003) and appropriately customized. See
Appendix B. We turn now from the discussion of participant groups to that of the materials set.

To assess the phonetic setting through the proxy of the CVE, a materials set was required
that would allow comparison of vowels in contexts that contrasted for voicing, with parallel
environments in the two languages. Clearly, actually occurring and familiar words are generally
to be preferred, because these are more likely than nonwords to evoke natural and fluent
pronunciations with naïve participants. The design of the materials set sought a broad range of
exemplars, but was necessarily constrained by the characteristics of the lexicon of each language.
For example, among words that were likely to be familiar, the prototypical word structure for
English would be monosyllabic, e.g., tube.¹⁷ Italian has an inventory of monosyllabic words, but

¹⁷ Greenberg (1997) looked at Switchboard, a corpus comprised of unrehearsed, unscripted, informal telephone
conversations, and found all of the thirty most common words to be monosyllabic and only ten of the one hundred
these tend to be function words, e.g., *la, tu, da, con*. Therefore, the default in Italian would be disyllabic, with most words in the language being subject to inflection, e.g., *tubo* [tu.bo].

Within each language, materials lists were constructed utilizing minimal or near-minimal pairs in which the vowel of interest was always stress-bearing, and the voicing contrast fell in a post-vocalic obstruent, e.g., *rib, rip*. This critical post-vocalic consonant or CVE-trigger, as it may be called, occupied coda position in the monosyllabic English cases where it created a closed syllable for the target vowel. However, it occupied onset position in the (minimally) disyllabic Italian cases, and left the target vowel in an open syllable. Coda stops are highly marked in Italian, being primarily restricted to cases of gemination in word-internal position, and to loan words in word-final position (Krämer 2009), ruling out a cross-language match in syllable structure. We attempted to counter this structural mismatch by adding (minimally) disyllabic English word pairs; these were divided into non-coronal, e.g., *stable, staple*, and coronal cases, e.g., *medal, metal*, the latter usually being subject to an English-specific phonological alternation, “flapping.”

To achieve near-equivalency despite the obvious cross-linguistic differences in syllable structure and default syllabic potential noted above, materials were kept as logically close as possible by sampling in equivalent ways in the two languages. For the critical target vowels, this meant incorporating into the design distinctions of height (high, mid, low) and backness (front, back), which are salient in both Italian and English, and bear on intrinsic duration. Despite this, inventory differences restrict the extent to which any cross-language parallelism can be achieved. On the one hand, the Italian vowel inventory is notably smaller than that of English, and subject to a more restricted distribution of the tense/lax distinction, where such distinction occurs only most common words in the corpus to be polysyllabic. According to his estimates, monosyllabic words represented some 22% of the lexicon in Switchboard yet accounted for just over 81% of the usage.
with mid vowels. On the other hand, many vowels of English, but not Italian, are routinely
diphthongized, wherein the vowel target changes within a single syllable.

For consonants, equivalency in sampling meant that materials pairs used obstruents that
are common to both languages, i.e., \(p\sim b, t\sim d, k\sim g\). The alveolar fricatives \(s\sim z\), which both
languages share, were excluded based on their manifestation in Italian. Orthographic \(s\) can be
realized as either [s] or [z], and systematically undergoes intervocalic voicing in northern
varieties of Italian. Moreover, orthographic \(z\) is realized as an affricate (Maiden & Robustelli
2000, Krämer 2009, among others). Details of the experimental materials are spelled out in
§2.2.2.

Data collection entailed recording each participant reading aloud short sentences, with
each sentence containing a single target word embedded in a carrier phrase. A carrier-phrase
production task was the chosen method as it was more likely to encourage natural speech as
compared with other elicitation protocols (such as reading a bare word list), and less likely to
distort critical vowel durations via processes such as utterance-final lengthening. To distract
attention from the relative uniformity that the materials design placed on the target word sets,
assorted fillers were intermingled and also presented to participants for recording, e.g., English
\textit{pair}, Italian \textit{bella}.

Following the recording sessions, the next step was measuring the vowel duration for
each utterance via segmentation and annotation of participants’ recorded utterances to isolate the
target vowel and the CVE-triggering consonant, e.g., [ɪ] and [p] in \textit{rip}. While listening to
recordings and simultaneously examining the corresponding waveform and spectrographic
representations, the investigator inserted segment boundaries and entered transcriptions for each
utterance, using the shape and intensity of the waveform and spectrogram as well as the presence of formants to make boundary decisions.

We segmented utterances into separate sections, from the beginning of the carrier phrase through to the completion of the target word. Utterances with disfluencies or mispronunciations in the immediate environment of the target vowel were marked as such (and subsequently discarded), as were those for which recording failed in some way. In order to prepare the vowel duration values for analysis following segmentation and annotation, duration measurements were extracted and exported to create data matrices over which statistical analyses (subject-based and item-based) could be conducted. Prior to analysis, we looked for outliers in the data. Taking into account that each speaker has a personal rate of speech and that rates of speech — and therefore durations — vary greatly across speakers (Klatt 1976), outlier identification was conducted for each speaker separately. Cross-checking the duration values with the items produced revealed that values could be exceptionally high — or, in some cases, exceptionally low — for given words. To include these values while preventing their distorting the overall data, a speaker-based limiting value was set rather than mandating any absolute cut-off, as detailed in §2.2.5.

2.2 Project implementation detail

The discussion in the preceding section provided the rationale for the project design. Decisions were shown to be motivated by such factors as the availability of speakers and the constraints of each language. The discussion turns now to the particulars, with extensive detail of the project execution described.
2.2.1 Participants: Foreign-born and U.S.-born

Participants in the foreign-born group (N=7, 4 of whom were male, and 3 female) ranged in age from 57 to 71 years (mean 62.3; standard deviation 6.3); their length of residence in an English-speaking environment ranged from 33 to 54 years (mean 41.0; standard deviation 7.3). Three additional participants who had been recruited into the foreign-born group were set aside as they fell outside the typical age range, did not have roots in or near Quaglietta, or did not meet the criterion of length of residence outside of Italy; a fourth additional participant had provided data for Italian only, and was similarly set aside.

Participants in the U.S.-born group (N=6, 2 of whom were male, and 4 female) ranged in age from 33 to 56 years (mean 42.8; standard deviation 8.6). One participant, born in London, had moved to the U.S. at five years of age. Two additional participants who were originally assigned to the U.S.-born group were set aside when they were found to have been born in Italy, and to have relocated to the U.S. at ages 9 and 10. A third additional participant, aged 87, was set aside on the basis of age, given the large gap (30+ years) between her and the next oldest participant in the group. To keep group sizes more nearly equivalent, one participant was included despite a non-Quagliaettan background; earliest exposure to Italian for this participant came via family with roots in Southern Italy.

The U.S.-born group was naturally younger, with no overlap in age range with the foreign-born group. However, the mean time spent in an English-speaking environment was very similar for the two groups, with almost complete overlap in the range (42.8 for U.S.-born, cf. 41.7 for foreign-born). Participants were recruited from an Italian-American social club located in the borough of Queens in New York City. The club brings together those with a

---

18 One participant volunteered age information only as “40+.” As it is certain that this participant falls into the age range for the U.S.-born group, 45 years was entered as a compromise figure for her age.
familial connection to Quaglietta and who live primarily in the tri-state area of New York, New
Jersey and Connecticut, with a population concentration in the immediate vicinity of the
clubhouse, where many settled upon their arrival from Italy. The clubhouse serves as a nexus of
activity, hosting such events as monthly meetings and celebrations associated with traditional
festivals, both religious and secular.

### 2.2.2 Materials design

The core set of Italian stimuli included 36 words — 18 minimal or near-minimal pairs —
representing 12 words (6 pairs) for each level of vowel height. The corresponding English
stimuli included an analogous set of 36 words. A further set of 24 multisyllabic English words
was added, representing 6 non-coronal pairs (a no-flap control set), and 6 coronal pairs (potential
flapping cases). Each word appeared twice in the experiment. A full listing of the materials is
offered in Appendix C for Italian and Appendix D for English, with examples in Table 2.1.

<table>
<thead>
<tr>
<th>Table 2.1</th>
<th>Illustrative stimulus word pairs, categorized by target vowel height and place of articulation of CVE-triggering consonant.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Italian, N=36</strong></td>
<td>Height</td>
</tr>
<tr>
<td>tipo</td>
<td>tubo</td>
</tr>
<tr>
<td>metodo</td>
<td>medico</td>
</tr>
<tr>
<td>grata</td>
<td>grado</td>
</tr>
<tr>
<td><strong>English, monosyllabic, N=36</strong></td>
<td></td>
</tr>
<tr>
<td>pick</td>
<td>pig</td>
</tr>
<tr>
<td>bet</td>
<td>bed</td>
</tr>
<tr>
<td>cap</td>
<td>cab</td>
</tr>
<tr>
<td><strong>English, multisyllabic, N=24</strong></td>
<td></td>
</tr>
<tr>
<td>staple</td>
<td>stable</td>
</tr>
<tr>
<td>metal</td>
<td>medal</td>
</tr>
</tbody>
</table>
One of the originally constructed Italian word pairs, *acropoli ~ acrobata*, proved excessively difficult for both participant groups. While numerous participants successfully produced *acropoli*, almost all failed to pronounce *acrobata* with appropriately placed stress, i.e., *acróbata*, and instead positioned stress after the model of the English counterpart to produce the incorrect form *acrobáta*. The investigator discarded this word pair and substituted an appropriate pair, taken from the fillers for this experiment, in its place.

The set of fillers for the experiment included 84 words for the Italian and 72 words for the English. Similar to the core set of data, the set of fillers was composed of words that would be familiar to the speakers yet with a greater variety of syllable structure and phoneme inventory, e.g., English *face, picnic, signature, van*, and Italian *bella, cane, festa, latte*. These fillers additionally served the purpose of providing additional data for later study, as discussed in the future research section.

2.2.3 **Elicitation protocol**

Target words were embedded in carrier phrases: for English, that phrase was *Say the word « ___ » to me*, and for Italian, *Dico la parola « ___ » di nuovo*. The consonant-initial phrase *to me* was deliberately selected for English to prevent re-syllabification, in the case of the monosyllabic stimuli, of the target word coda consonant with the following segment (as may occur, for example, with the vowel-initial word *again*). For consistency, a consonant-initial phrase, *di nuovo*, was also utilized in Italian. Stimulus sentences were presented one at a time in written form on a computer screen in a comfortably large font, and participants were instructed to read each of these aloud. There was no time pressure for the initiation of utterances, and participants themselves chose the pace with which they proceeded from one stimulus to the next.
Target words and intermingled fillers were pseudo-randomized for each participant, separately, to ensure that members of a minimal pair rarely, if ever, occurred consecutively.

Sessions were conducted on different days, one for each language, to keep the recording sessions as two separate events, to the extent possible. The first session was always in Italian, and the second in English. The investigator presented instructions and communicated in the language corresponding to the task, switching languages as necessary. The session began with verbal instructions followed by three written lines shown consecutively on the monitor, in separate screen shots:

Say the sentence shown on the screen.
Please remember to speak naturally.
Here are some sentences for practice.

Deliberately chosen practice items included such words as *wrap, eagle* and *badge* in English and *strada, oca* and *società* in Italian. Halfway through the stimulus list, a message on the monitor indicated the midway point had been reached, and suggested the speaker take a break. Each experimental item and filler was presented twice during the recording session, once in the first half of the stimulus list and again in the second half. Sessions typically lasted under an hour, with an average of 20-30 minutes spent recording and 15 minutes spent on the questionnaire, usually administered during the first of the two sessions. The questionnaire, written in English, was completed only after the recording session had concluded. Sessions took place at the Quagliettana Society clubhouse in Astoria, Queens, at the Linguistics Program of the CUNY Graduate Center, or in a private home/office space, as chosen by the participant.

Stimulus sentences were presented using DMDX (Forster & Forster 2003), a program for stimulus display and data capture. Participants initiated the display and began the recording with the push of a button on the keyboard. Speakers read the stimuli aloud into a Logitech noise-
cancelling adjustable headset microphone connected to a computer. The boom microphone had a
frequency response from 100 to 10,000 Hz. Each individual display timed out 20 seconds after
initiation, but was typically terminated earlier, when the participant pressed a button to advance
to the following utterance. DMDX auto-labeled each waveform, uniquely identifying the sound
file for speaker, language, item and token. Files were saved in waveform audio format (.wav).

2.2.4 Utterance segmentation

The number of utterances per participant was 192 in total. The main portion, 144 of the
utterances, comprised the core data for a given participant: 72 Italian language stimuli (36
targets, each presented twice), and likewise 72 English language stimuli using monosyllabic
targets. The remaining 48 utterances were English language stimuli using multisyllabic targets
(24 targets, each presented twice). Utilizing a waveform editing program (SIL International’s
Speech Analyzer, 2007), the investigator segmented each utterance into 7 mutually exclusive
regions (6 in the case of English monosyllables) and subsequently entered the corresponding
phonetic transcription inside the file. Transcriptions were especially useful for identifying and
categorizing speech errors (see §2.2.5) within the utterances. Segmentation was exhaustive,
beginning at the onset of the carrier phrase and running through the end of the target word. The
remaining part of the carrier phrase, the portion following the critical target word, was not
segmented. The regions were as follows:
Carrying phrase beginning, i.e., English *say the wor*; Italian *dico la*

Carrying phrase ending, i.e., English *di*; Italian *parola*, including any following pause plus any closure for target-initial consonant

*Region 3*  
*C1*, phonation of consonant(s) preceding target vowel

*Region 4*  
*V1*, target vowel

*Region 5*  
*Closure*, closure for post-vocalic consonant

*Region 6*  
*C2*, phonation of consonant following target vowel

*Region 7*  
*V2...#*, target word remainder for utterances with multisyllabic targets, e.g., English *metal*, *cabinet*; Italian *moto*, *tipo*

Regions are depicted by the illustrative schematic offered by Figure 2.2:

Figure 2.2  
Illustration of spectrogram-supported waveform segmentation and transcription for the utterance, *Dico la parola « tipo » di nuovo*.

Region definitions took advantage of carefully chosen and reliably identifiable acoustic landmarks. That is, the segmentation plan allowed capture of the relevant data for analysis while affording consistency from speaker to speaker, and from case to case. Segment boundaries for
each waveform were placed with support from spectrographic evidence and playback of the audio recordings. Segmentation of Region 4 necessitated examination of the two displays in tandem: boundary decisions were informed by the presence of the upper formants of the target vowel — as found within the spectrogram — in conjunction with the periodic waveform. Spectrographic evidence also helped identify the noise burst associated with voiceless stops, as seen in Regions 3 and 6 of Figure 2.2. Consistency in boundary decisions across speakers and items played a crucial role in the segmentation process, illustrated with the following example. The core data contained a substantial number of $r$-initial stimuli wherein the C1 and the V1 overlapped. As such, locating a boundary between $r$ and the beginning of the vowel was not possible. Priority was given to identifying the vowel boundaries while additionally confirming the presence of an $r$-segment. Any additional ambiguous cases with other C1s were handled analogously.

The first portion of the utterance, Region 1, was designed as a rate estimator for a given participant. Speakers produced this portion of the utterance fluidly, with an absence of variable pausing. Region 2 commonly held — for all speakers — pauses of varying length immediately preceding the target word, and thus measurements for this region were not utilized for analysis due to their irregularity across speakers and cases. For the English, Region 1 encompassed the area from the start of the fricative for *say* to the end of major vocalization for *word*, easily located given the closure that preceded the $d$. Region 2 began with that closure period and ended immediately preceding phonation for the initial consonant(s) of the target word. Phonation variants associated with these two consonantal regions included frication, sonorant resonance, release burst and aspiration. For the Italian, Region 1 started at the plosive burst for *dico la* [dikola] and ended with the closure period preceding *parola* [parola]. Region 2 began with
*parola* and ended immediately preceding phonation for the initial consonant(s) of the target word.

Regions 3 through 6, highlighted in Figure 2.2, comprised the four most relevant regions, i.e., the critical portion of the target word. Any prevoicing associated with the C1 (Region 3) or the C2 (Region 6) was included in the closure preceding the respective consonant. Monosyllabic cases terminated with Region 6, while Region 7 contained the word remainder for multisyllabic cases. Content for Region 7 ranged in segment length, e.g., a single segment in *tipo* and multiple segments in *cabinet*, making measurements for this region inherently irregular. Having this full suite of regions enables reconstruction of the duration of critical parts of the utterance. To illustrate, the combination of Regions 3 through 7 (or Region 6 in the case of monosyllabic words) recreates the target word. Smaller sections can also be recreated, e.g., the closure can be associated with V1 or with C2 (or both) depending upon the desired analysis.

### 2.2.5 Data extraction and filtering

Following segmentation, duration measurements were extracted from Speech Analyzer and subsequently scrutinized for “usability.” As expected, not all utterances afforded data that could be used for the intended analysis. Occasionally, the elicitation protocol failed, leading to a data-capture error, i.e., missing data, such as when a participant mistakenly touched the keyboard to advance to the following utterance before or in the process of speaking, leaving no time for the utterance to be satisfactorily recorded. More commonly, however, unusable data points resulted from a production error in the critical target sequence, i.e., Regions 3 through 6. Patterns in the resulting categorization for usability are summarized in Table 2.2 below for Region 4 data,
which are both central to the aims of the project and typical of the result that emerged across regions.19

Table 2.2  Mean number of vowel-duration measurements deemed usable, per participant, as a function of stimulus type. Percentages cast values for usability relative to the maximum possible.

<table>
<thead>
<tr>
<th></th>
<th>Usable</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Foreign-born participants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian (N=72)</td>
<td>68.1</td>
<td>94.6%</td>
</tr>
<tr>
<td>English, monosyllabic (N=72)</td>
<td>55.0</td>
<td>76.4%</td>
</tr>
<tr>
<td>English, multisyllabic (N=48)</td>
<td>38.7</td>
<td>80.7%</td>
</tr>
<tr>
<td><strong>U.S.-born participants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian (N=72)</td>
<td>60.3</td>
<td>83.8%</td>
</tr>
<tr>
<td>English, monosyllabic (N=72)</td>
<td>68.7</td>
<td>95.4%</td>
</tr>
<tr>
<td>English, multisyllabic (N=48)</td>
<td>46.8</td>
<td>97.6%</td>
</tr>
</tbody>
</table>

Language of education is apparently the major determinant of successful performance of “read speech” in the elicitation protocol. That is, foreign-born participants showed notably lower rates of usable data for English language than for Italian language stimuli, while U.S.-born participants showed the reverse pattern. The somewhat lower usability rates found with foreign-born/English (76.4%, 80.7%) versus U.S.-born/Italian (83.8%) might be attributable to transparency differences in the orthography of these languages — Italian having a more transparent orthography than English — despite the familiarity of the target words involved.

A more detailed evaluation of the sources of unusability seems to indicate that the data loss often stemmed from lack of familiarity of the written form of the target word, or competition

19 The only departure from the values in Table 2.2 across Regions 3 to 6 arose in Region 6 (post-vocalic consonant); these departures were seen expressly in the case of English monosyllables. Participants occasionally triggered a premature halt to recording, with the result that the consonant was not fully captured. In such cases, the quality of the C2 was clear, but the boundaries were not. Data loss for this region was slightly greater for foreign-born participants (73.4% usable) and considerably greater for U.S.-born participants (87.3% usable) as compared with the other regions.
with a more familiar word. A representative example comes from the witty ~ widow materials pair, which fared poorly with the foreign-born group. Participants frequently produced white in lieu of witty, wherein [wɪ.ti] became [wæt], a change in both the vowel quality and the syllable structure. The word widow simply became window, an example of an orthographically similar and more familiar word emerging. Target vowel production errors by foreign-born speakers with the monosyllabic English language stimuli also exemplified this pattern, where peck [pɛk] was read as pick [pɪk], peg [pɛg] as pig [pɪg], and nod [næd] as node [nɒd].

The foreign-born group also tended toward errors of syllabicity: vowel epenthesis in post-C2 position accounted for roughly one-third of the errors found within the monosyllabic English stimuli. The ensuing resyllabification, which prevented a closed syllable, e.g., seed [sid] to [si.də]; mad [mæd] to [mæ.də], changed the environment for the critical vowel-consonant sequence from tautosyllabic to heterosyllabic, making the data unusable. Syllabicity errors also went in the opposite direction, with suffixal vowels occasionally deleted, such that a multisyllabic word was restructured into a monosyllable. The data again became unusable, in this case because the environment for the critical vowel-consonant sequence changed from heterosyllabic to tautosyllabic, e.g., lucky [lʌ.ki] read as luck [lʌk].

With the U.S.-born speakers, errors appeared to have the same sources — lack of familiarity of the written form of the target word or competition with a more familiar word — but varied more in their manifestation. At times, U.S.-born speakers gave an English-language orthographic interpretation of an Italian vowel letter, e.g., Italian lìte [li.te] ‘quarrel’, a less-familiar word, produced as [lɐt.te]. Stress errors by this group with the Italian tri-syllabic stimuli were common, where stress erroneously fell on the penultimate syllable of a tri-syllabic Italian
word, rather than the antepenultimate, e.g., *metodo* ['mɛ todo] as *[mɛ todo]; *cupola* ['kupola] as *[ku'pola].

In addition to scrutinizing the data for production errors, as detailed above, and removing cases in which the speaker failed to achieve the target sequence, it was also necessary to identify outliers in the resulting set. Outliers were defined as values that fell outside the typical range for the data set. To prevent outright exclusion of certain data on the basis of intrinsic duration (e.g., prevocalic fricatives commonly correspond to greater duration as compared with oral stops), a speaker-based limiting value was implemented. As detailed below, these cases were adjusted based on standard practice to minimize the impact of extreme values in the data.

As a final step prior to data analysis, usable duration values were reviewed separately for each region and for each speaker within a given stimulus set, e.g., Region 4 (target vowel) data for a given participant, for Italian. For each such data subset, upper and lower limiting values were calculated as the mean ± 2 standard deviations. Values falling beyond these limits (outliers) were replaced by the corresponding limiting value. No distinctions were made with respect to categorizations within the stimulus set as defined in the materials design, e.g., target vowel height distinctions and C2 voicing contrasts.

The average number of outliers for Region 4 data, per participant, is summarized in Table 2.3 below. Note that “usable” data values, included here for reference, are carried over from Table 2.2.
Table 2.3  Mean number of vowel-duration outliers, per participant, as a function of stimulus type. Percentages indicate values based on usable data.

<table>
<thead>
<tr>
<th></th>
<th>Usable</th>
<th>Outliers</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Foreign-born participants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>68.1</td>
<td>1.4</td>
<td>2.1%</td>
</tr>
<tr>
<td>English, monosyllabic</td>
<td>55.0</td>
<td>1.9</td>
<td>3.4%</td>
</tr>
<tr>
<td>English, multisyllabic</td>
<td>38.7</td>
<td>2.1</td>
<td>5.5%</td>
</tr>
<tr>
<td><strong>U.S.-born participants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>60.3</td>
<td>2.3</td>
<td>3.9%</td>
</tr>
<tr>
<td>English, monosyllabic</td>
<td>68.7</td>
<td>2.5</td>
<td>3.6%</td>
</tr>
<tr>
<td>English, multisyllabic</td>
<td>46.8</td>
<td>1.0</td>
<td>2.1%</td>
</tr>
</tbody>
</table>

Percentages fell within the expected range for outliers in the data set. Data are positively skewed, with intrinsically longer vowels, diphthongized vowels and those followed by a voiced segment most likely to fall in the upper tail. The approach applied to outliers limited the extent to which these extreme values could distort the data while not discarding any data. The listing of outliers per participant for all relevant regions can be found in Appendix E.

2.2.6 Considerations shaping the data analysis

To compare duration values in the voiced and voiceless context for each language, the chosen statistical model was analysis of variance (ANOVA). In preparing data for analysis, mean duration values were computed over usable data, as described in the preceding section, for each language set; unusable data will be referred to as “missing” data. Each item had maximally two tokens per speaker. Means were averaged over tokens to minimize the impact of missing data.

Analyses were conducted using means for each speaker — collapsing over items — for the subject-based analyses, and means for each member of a word pair — collapsing over
speakers — for the item-based analyses. Values were calculated separately for each region of the utterance (as described in §2.2.4) based on [voice] value, i.e., trigger consonant voicing. While analyses were conducted over absolute durations, these values can be subject to speaker-specific speech rate variation as well as item-specific differences in the phonetic environment. Ratios are a desirable form of data description, as they help abstract away from these sources of potential differences; but their sensitivity to missing data made calculation of a true ratio for each word pair untenable for this data set. Instead, a derived ratio was determined by calculating the ratio for each speaker, i.e., dividing the mean duration value for the speaker in the voiced setting by the mean duration value in the voiceless setting, then averaging over those speaker-based ratios; likewise, for items and item-based ratios. This derived ratio was useful for the purposes of discussion and comparison with prior studies. The discourse now shifts from how the study was conducted to the findings of the study. Results are reported, along with a discussion of the data patterns therein.
Chapter 3: Data patterns and discussion

The previous chapter described the methodology for collecting and preparing the data for study. The present chapter presents the results and analyses, describing how two groups of bilingual speakers employ the consonantal voicing effect in each of their languages. Data patterns are first presented for the core stimuli, and are immediately followed by an exploration of speaker-based variability found with those data. The core cases are followed by cases matched for syllable structure, leading into a discussion of how each speaker group manages syllabic and timing patterns in the two languages. Following a glimpse at the behavior found in the particular case of flapping, the chapter closes with a description of the data patterns for the closure region, for both stimulus sets.

3.1 Consonantal voicing effect: Core cases

Vowel duration values for the core cases were first analyzed using a three-way design having factors Speaker Group, Language of Utterance and [Voice] Value. Speaker Group refers to the 7 participants for the foreign-born group, on the one hand, and the 6 participants for the U.S.-born group, on the other. Language of Utterance refers to the 18 minimal pairs for the Italian stimulus set (e.g., tipo, tubo) as uttered within the carrier phrase Dico la parola « ___ » di nuovo, cf. the 18 minimal pairs for the English stimulus set (e.g., pick, pig) as uttered within the carrier phrase Say the word « ___ » to me. [Voice] Value refers to the immediate post-vocalic stop environment having values voiceless and voiced. The consonantal voicing effect (CVE) is expressed in terms of the difference in absolute duration between the voiced and voiceless environments.
Vowel duration patterns for both speaker groups are summarized in Figure 3.1, below. Unless otherwise noted, values are drawn from item-based analyses.

![Figure 3.1](image)

**Figure 3.1** Mean duration (and standard error) of critical vowel for foreign- and U.S.-born speaker groups for the core cases, as a function of language of utterance and [voice] value in the post-vocalic environment.

As suggested by the overall data patterns seen in Figure 3.1, the omnibus analysis of variance revealed no three-way interaction of speaker group, language of utterance and [voice] value, $F_{1}<1, F_{2}(1,34) = 1.17, p>.25$. There were, however, two significant two-way interactions. [Voice] value interacted with speaker group, $F_{1}(1,11) = 4.76, p=.052, F_{2}(1,34) = 81.88, p<.001$, and also with language of utterance, $F_{1}(1,11) = 12.00, p<.01, F_{2}(1,34) = 36.48, p<.001$. The voice by group interaction can be seen to arise because foreign-born speakers consistently exhibited smaller CVE magnitudes than did U.S.-born speakers (see Table 3.1, below, for CVEs). The voice by language interaction reflects the fact that, for both speaker groups, CVEs were smaller for Italian as the language of utterance than for English as the language of utterance. In both of these interactions, the role of the [voice] value factor is illustrated via the
magnitude of the CVE. There was no reliable interaction of language of utterance and speaker
group, $F_1(1,11) = 2.03, p>.10, F_2(1,34) = 15.07, p<.001$.\textsuperscript{20}

We now look in fuller detail at the patterns within each speaker group, and consider first
data for foreign-born participants, presented in Figure 3.1’s left panel. Despite its relatively
small magnitude, subanalyses found the CVE to be reliable for Italian utterances, $F_1(1,6) =
13.74, p<.05, F_2(1,17) = 28.38, p<.001$. The CVE was also reliable for English utterances,
$F_1(1,6) = 7.28, p<.05, F_2(1,17) = 120.05, p<.001$, which exhibited a larger magnitude.

A similar pattern was found for U.S.-born participants, presented in Figure 3.1’s right
panel. Here, however, the magnitude of the effect was even greater, for both languages of
utterance. Subanalyses confirmed reliability of the CVE in each instance: for Italian, $F_1(1,5) =
30.29, p<.01, F_2(1,17) = 224.18, p<.001$, and for English, $F_1(1,5) = 70.70, p<.001, F_2(1,17) =
125.66, p<.001$. Once again, the magnitude of the CVE can be seen to be greater when English
was the language of utterance.

Table 3.1 lists values for the CVE, for both speaker groups.

<table>
<thead>
<tr>
<th></th>
<th>Foreign-born</th>
<th>U.S.-born</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVE-Italian</td>
<td>16.8 (3.1)</td>
<td>49.3 (3.3)</td>
</tr>
<tr>
<td>CVE-English</td>
<td>60.9 (5.6)</td>
<td>86.5 (7.7)</td>
</tr>
</tbody>
</table>

The consonantal voicing effect is known to be present cross-linguistically, and the size of
the basic effect is usually seen as consistent with a phonetic process. The greater magnitude

\textsuperscript{20} Despite reaching significance in the item-based analysis, the interaction is treated as unreliable due to the lack of
significance — or even trend — in the subject-based analysis.
found with English as compared with other languages (see §1.1.1 and §1.1.2), including Italian, is considered to be a phonological enhancement of a basic phonetic process (see §1.1.4). Each group exhibited CVE magnitudes in the native or dominant language — Italian for foreign-born speakers, English for U.S.-born speakers — that were consistent with values found for monolingual speakers in earlier studies. In the second or non-dominant language, however, values for each group differed from those found for monolingual speakers, manifesting a compromise setting, i.e., a value in between the established language-specific timing settings for Italian and English.

Plausibly, the path for arrival at that compromise setting differs between groups. On the entirely natural assumption that each speaker group takes as a starting point the setting consistent with the native (or dominant) language, compromise setting would be arrived at via adjustments involving implementation, on the one hand, but suppression, on the other, of the phonological component. The hypothesized routes to compromise are summarized in Table 3.2 below and further described for each speaker group following the table.

**Table 3.2** Summary of hypothesized CVE mechanisms for each speaker group, as a function of language of utterance.

<table>
<thead>
<tr>
<th></th>
<th>Foreign-born</th>
<th>U.S.-born</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CVE-Italian</strong></td>
<td>NATIVE LANGUAGE</td>
<td>NON-DOMINANT LANGUAGE</td>
</tr>
<tr>
<td></td>
<td>Phonetic effect only</td>
<td>Phonetic effect + partially suppressed phonological effect</td>
</tr>
<tr>
<td><strong>CVE-English</strong></td>
<td>SECOND LANGUAGE</td>
<td>DOMINANT LANGUAGE</td>
</tr>
<tr>
<td></td>
<td>Phonetic effect + partially implemented phonological effect</td>
<td>Phonetic effect + phonological effect</td>
</tr>
</tbody>
</table>
The CVE-Italian for foreign-born speakers, equal to a derived ratio of 1.10, appears to be consistent with the presence of a purely phonetic effect as would be found with monolingual speakers of Italian. The data suggest that foreign-born speakers maintained this language-specific setting. This CVE finding can be compared with the findings of Esposito (2002), who found a greater duration difference with native speakers of Italian than was found in the present study (28 ms vs. 17 ms), consistent with a 1.23 ratio. It is possible that the duration difference across studies was due to the difference in materials, i.e., a matter of phonetic context. In the Esposito study, the target vowel was preceded and followed by the same consonant (e.g. abab). It may be that comparing vowels surrounded by voiced consonants with vowels surrounded by voiceless consonants increased the magnitude of the duration difference. In their 2001 study of English, Hillenbrand et al. showed the impact of the [voice] value of the prevocalic consonant, where vowel durations were greater on the order of 20-40 ms when the preceding consonant was voiced as compared with voiceless. The stimuli used by Esposito could therefore have amplified the difference in voicing environments. We understand the duration difference found across studies for Italian as the language of utterance to be likely due to the differences in the stimulus set, i.e., the phonetic environment.

The CVE-English for the foreign-born speakers fell just below the range of what has routinely been found with English for native speakers, where vowel durations were typically at least 50% greater in the voiced context (see Table 1.3 in §1.1.1). The 1.45 ratio found for the L2 English was simultaneously smaller than the expected value for monolingual speakers of English, and greater than the expected value for monolingual speakers of Italian (even against the 1.23 ratio reported by Esposito). The sizable CVE suggests that foreign-born speakers start with the basic phonetic component, as found with the L1 Italian, and then manage a partial
implementation of the phonological mechanism consistent with the L2 English. It appears that participants, following an extended period of exposure to English, have both perceived the greater magnitude of the consequences of the voiced-voiceless contrast in English as compared with Italian, and at least partially incorporated that in their production of English.

For the U.S.-born speakers, the large CVE-Italian, equal to a derived ratio of 1.38, indicates that that the U.S.-born speakers who grew up with exposure to both languages did not maintain a strictly phonetic setting for the Italian. These speakers appear to start with the basic phonetic process and the phonological enhancement consistent with the dominant language of English. To arrive at the compromise setting in the L2 Italian, speakers partially suppress the sizable enhancement consistent with the English language-based phonological mechanism.21

The greater CVE-English for the U.S.-born group, equal to a derived ratio of 1.66, was in line with the findings of other English-language studies, and was especially similar to the findings of Chen, who found a ratio of 1.63 (92 ms). The size of the CVE in English for U.S.-born speakers seems to indicate the expected phonetic effect in combination with a phonological enhancement. Table 3.3 lists the ratios for each language, for both speaker groups.

<table>
<thead>
<tr>
<th>Table 3.3</th>
<th>Voiced-to-voiceless ratios calculated for foreign- and U.S.-born speaker groups for the core cases, based on language of utterance.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Foreign-born</td>
</tr>
<tr>
<td>Italian ratio</td>
<td>1.10</td>
</tr>
<tr>
<td>English ratio</td>
<td>1.45</td>
</tr>
</tbody>
</table>

21 As detailed in this section, each speaker group exhibited a compromise setting in the L2 (the non-dominant language). Analyses were conducted to compare vowel duration values for each speaker group in the respective L2 using a two-by-two design having factors Case and [Voice] Value, where case was English for the foreign-born group and Italian for the U.S.-born group. The interaction terms of the analyses offer no evidence of a difference between the compromise settings adopted by each group in their L2. Logically, the groups could arrive at settings that are similar or distinct from one another, but a study with substantially increased power is required to establish the detail on this point.
In summary, CVE effects vary as each group differs in their timing setting for each language. Foreign-born speakers exhibited the presence of the CVE in both of their languages, but with different magnitudes. The size of the CVE was much smaller with Italian as the language of utterance, and was comparable to the expected size for monolingual speakers of Italian; the size of the CVE was much greater when English was the language of utterance, approaching — but not fully reaching — the expected size for monolingual speakers of English. The U.S.-born speakers also exhibited the CVE in both of their languages and with the same general pattern as the other group: smaller for the Italian and greater for the English. This group differs from the other group with respect to the magnitude of the effect: the CVE-Italian was larger than would be expected for monolingual speakers of Italian, and the CVE-English was as expected for monolingual speakers of English. For a better understanding of the variation within this study, we turn now to a consideration of speech rates.

3.1.1 Speaker speech rates

Rate variation arises from idiosyncratic speaker differences, e.g., being naturally fast or slow, as well as factors such as a speaker’s emotional state, the level of formality of the discourse, and, in the present case, the demeanor of a participant in response to the experimental setting. Usage of the CVE and the derived ratio help abstract away from these differences as well as provide a means for comparison of language-specific data patterns both within this study and with prior studies. An exploration of speech rates and variation found within the present study can provide insight into the participant groups as well as individual speaker differences. The beginning portion of the carrier phrase was designated the rate estimator (see §2.2.4), a region where the speech rate was expected to be neutral with respect to the [voice] value of the
post-vocalic consonant. The phrases were somewhat parallel across the language of utterance, composed of approximately three syllables in each language, but the content differed phonetically: Italian: [di.ko.la]; English: [seɪ.ðə.wə].²² Values were analyzed using a three-way design with factors Speaker Group, Language of Utterance and [Voice] Value.

Rate estimator region duration patterns for both speaker groups are summarized in Figure 3.2, below. Values for this figure are drawn from the subject-based analyses.

---

Figure 3.2  Mean duration (and standard error) of rate estimator region for foreign- and U.S.-born speaker groups, as a function of language of utterance and [voice] value in the post-vocalic environment.

The omnibus analysis of variance revealed no three-way interaction of speaker group, language of utterance and [voice] value, p>.25 in both instances. As expected, [voice] value did not interact with language of utterance or with speaker group, p<.10 in all instances. There was one significant two-way interaction. Speaker group interacted with language of utterance, $F_1(1,11) = 10.06$, p<.01, $F_2(1,34) = 232.69$, p<.001, reflecting the duration difference between speaker

²² While the final /d/ of word was typically present in the utterance, the investigator chose to terminate segmentation for this region at the end of major vocalization for the phrase, which immediately preceded the closure for the /d/. See §2.2.4 for further details of the segmentation.
groups when English was the language of utterance. The same analyses showed no reliable
effect of [voice] value of the trigger consonant for either group, Fs<1.

For the foreign-born group, subanalyses showed the mean duration to be markedly
greater with English as the language of utterance, $F_1(1,6) = 19.24$, $p<.01$, $F_2(1,34) = 411.36$, $p<.001$. For the U.S.-born group, the mean duration was somewhat greater numerically for
English as language of utterance than for Italian, but the effect was not reliable, having reached
significance only in the item-based analysis, $F_1(1,5) = 1.92$, $p>.10$, $F_2(1,34) = 35.05$, $p<.001$.
Means for the rate estimator region for both speaker groups, collapsed over the [voice] value of
the trigger consonant, are summarized below in Table 3.4.

<table>
<thead>
<tr>
<th></th>
<th>Foreign-born</th>
<th>U.S.-born</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italian</td>
<td>544 (59.1)</td>
<td>511 (18.4)</td>
</tr>
<tr>
<td>English</td>
<td>796 (81.0)</td>
<td>549 (34.4)</td>
</tr>
</tbody>
</table>

The rate estimator region preceded the target word. Any differences found for this region
ostensibly due to the [voice] value (numerically only, as no difference was found statistically)
would have been related to individual stimuli, not to voicing. For example, word familiarity
might impact speech rate, i.e., a less familiar target word may correspond with an overall slower
rate of speech. However, no systematic correlation between word familiarity and the [voice]
value of the trigger consonant inside the target word would be expected.

Durations for the rate estimator region were (at least numerically) greater for English
than for Italian, regardless of speaker group, suggesting a naturally longer phonetic composition
for the English. As indicated by the standard error values as displayed in Figure 3.2 (and listed
in Table 3.4), a greater amount of variability was found with the English than with the Italian, and a much greater amount of variability was found with the foreign-born group as compared with the U.S.-born group. We take duration in this region to represent the comfort level of participants in each language and with the task overall. Given a prototypical intrinsic duration for the rate estimator region, a greater degree of comfort would lead to a reduction of this standard duration while a lesser degree of comfort would lead instead to an augmentation. Performing a task in one’s language of education would likely correspond to a greater degree of comfort with these speaker groups. As such, if the language of utterance was the speaker’s language of education, the intrinsic duration would be expected to be minimized, i.e., shorten. Conversely, the intrinsic duration would be expected to lengthen if the language of utterance was not the speaker’s language of education. The speaker-based means for this region are graphed for both speaker groups in Figure 3.3, below. Means for each individual foreign-born speaker are presented in the left panel and, for each individual U.S.-born speaker, in the right panel.

Figure 3.3  Mean duration of the initial portion of the carrier phrase, the rate estimator, for each speaker, as a function of language of utterance. The triangular marker sets apart an especially slow speaker within the foreign-born group.
Italian was the language of education and the dominant language for the majority of the foreign-born speakers. They tended to have more experience and greater proficiency with the Italian language than did the U.S.-born speakers. They also tended to have fewer years of formal schooling than did the U.S.-born group. As immigrants, their social, professional and linguistic experiences were variable; that variation manifests itself in Figure 3.3. A much greater amount of variation was demonstrated for the Italian by the foreign-born speakers than by the U.S.-born speakers, even setting aside one speaker who exhibited an exceptionally slow rate of speech (identified in Figure 3.3 by the triangle-shaped marker).

English was the language of education and the dominant language for the U.S.-born speakers. In both languages, the U.S.-born participants clustered together much more closely than did the foreign-born participants, and tended to be faster overall. Being more at ease in the experimental setting was likely one of the elements contributing to the faster rate of speech for this group, regardless of language. Meanwhile, the slower rate of the foreign-born participants could be related to such factors as their more advanced age, making them naturally slower, and greater caution being exercised in the experimental setting.

The difference in rate estimator duration means across language was likely inflated for the foreign-born speakers, where the language of education was Italian but the intrinsically longer rate estimator carrier phrase was English. Conversely, the differences in means across language were likely reduced for the U.S.-born speakers, where the language of education and the intrinsically longer rate estimator carrier phrase were both English. If we were able to control for phonetic composition, we would expect to find the smallest mean for the rate estimator region with the U.S.-born speakers.
3.1.2 Intrinsic vowel duration

Another source of variability within the data is intrinsic vowel duration differences. These differences are evident with a height-based depiction of the data, as intrinsic vowel durations track roughly with height: low vowels tend to be of greater duration than non-low vowels (House & Fairbanks 1953, Fourakis 1991, among others).

For both speaker groups, vowel duration differences based on height distinctions were present in both languages. Appendix F provides a height-by-language mapping of the data with the core cases. The pattern of duration differences with respect to height found with the data were all as expected based on intrinsic vowel durations. Vowel height was built into the design to ensure breadth in the materials and systematic sampling across the two languages, but height distinctions were set aside in the analyses.

While the stimuli chosen for the core cases were specifically selected as default cases for each language, this meant that the English language stimuli were monosyllabic while the Italian language stimuli were (minimally) disyllabic. Given the possibility that differences in the CVE are attributable to syllabicity, the mismatch in syllable structure was addressed through the addition of multisyllabic English language stimuli. With the English monosyllabic stimuli, the trigger consonant for the consonantal voicing effect was in the same syllable as the target vowel, i.e., tautosyllabic. With the multisyllabic stimuli, the triggering consonant was instead in the syllable following the target vowel, i.e., heterosyllabic. We turn now to the results of this analysis.
3.2 Consonantal voicing effect: The heterosyllabic environment

As with the core cases, vowel duration values were first analyzed using a three-way design having factors Speaker Group, Language of Utterance and [Voice] Value. With this analysis, language of utterance refers to the 6 non-coronal minimal pairs for the English stimulus set (e.g., soccer, soggy) and 6 minimal pairs which were a subset of the Italian stimulus set (e.g., sacro, sagra). The 6-pair subset was a sampling from the 18-pair master Italian stimulus set, with items chosen to best match the English stimuli. Selection considerations included target vowel height and trigger consonant place of articulation. In all cases, the trigger consonant was heterosyllabic, positioned at the onset of the syllable following the target vowel (as compared with the core cases, where the trigger consonant was tautosyllabic for the English stimuli). Place of articulation for the English stimuli was exclusively non-coronal, as coronals would be subject to flapping in this context.

Vowel duration patterns for both speaker groups are summarized in Figure 3.4, below.

---

23 To compare the 18-pair set of Italian stimuli and the 6-pair subset, subject-based vowel duration values were analyzed for each participant group using a two-by-two design with factors Data Set and [Voice] Value. For each of the groups, there was no main effect of the data set and no interaction (foreign-born: ps>.25 for data set and interaction; U.S.-born: F<1 for the interaction with a strong trend for data set, p=.055). Values for the subset were considered to be representative of the full set of Italian stimuli. See Appendix G for the tables listing out the vowel duration means (and standard deviations) for both sets of Italian stimuli for each speaker group.
With these data, the omnibus analysis of variance revealed a three-way interaction of speaker group, language of utterance and [voice] value, $F_{1}(1,11) = 6.14, p<.05$, $F_{2}(1,10) = 6.26, p<.05$.

The interaction suggests that the magnitude of the CVE did not follow a consistent pattern across language of utterance for the two speaker groups.

Two-way subanalyses were conducted within each speaker group to explore the CVE patterns. We look first at the data for the foreign-born participants, presented in Figure 3.4’s left panel. Critical vowel duration was found to be greater when the trigger consonant was voiced rather than voiceless, $F_{1}(1,6) = 10.40, p<.05$, $F_{2}(1,10) = 14.51, p<.01$. Additionally, vowel duration was significantly greater when Italian was the language of utterance as compared with English, $F_{1}(1,6) = 33.71, p<.01$, $F_{2}(1,10) = 5.2, p<.05$. The CVEs were similar across language of utterance (16.8 ms, cf. 18.8 ms), $Fs < 1$ in both subject- and item-based analyses. See Table 3.5 for a summary of CVEs for both speaker groups.

**Figure 3.4** Mean duration (and standard error) of critical vowel for the multisyllabic English and a subset of the Italian stimuli for foreign- and U.S.-born speaker groups, as a function of language of utterance and [voice] value in the post-vocalic environment.
Data for the U.S.-born participants are presented in Figure 3.4’s right panel. With this speaker group, we found an interaction due to the difference in magnitude of the CVE as a function of the language of utterance, $F_1(1,5) = 9.61$, $p<.05$, $F_2(1,10) = 5.83$, $p<.05$. Further subanalyses confirmed the CVE to be reliable both for the Italian, $F_1(1,5) = 79.10$, $p<.001$, $F_2(1,5) = 59.54$, $p<.001$, and for the English, $F_1(1,5) = 44.71$, $p<.01$, $F_2(1,5) = 18.44$, $p<.01$. Contrary to the findings with the core cases, the CVE magnitude was greater for the U.S.-born speakers with Italian as the language of utterance than with English. Table 3.5 summarizes the CVEs for both speaker groups for the exclusively multisyllabic environment.

**Table 3.5**

Consonantal voicing effect (and standard error) in milliseconds for foreign- and U.S.-born speaker groups for the multisyllabic English and a subset of the Italian stimuli, as a function of language of utterance.

<table>
<thead>
<tr>
<th></th>
<th>Foreign-born</th>
<th>U.S.-born</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVE-Italian</td>
<td>16.8 (5.3)</td>
<td>46.7 (6.0)</td>
</tr>
<tr>
<td>CVE-English</td>
<td>18.8 (7.7)</td>
<td>26.0 (6.1)</td>
</tr>
</tbody>
</table>

The data patterns for the foreign-born speakers deviated from those seen with the core cases in that here the CVE did not differ significantly across language. Figure 3.4 illustrates an analogous pattern cross-linguistically for this group; in fact, the CVE in the two languages differed by only 2.0 ms (see Table 3.5). However, given that absolute durations were greater for the Italian than for the English, the derived ratios differed: 1.09 for Italian and 1.16 for English. As with the core cases, the CVE-Italian for the foreign-born speakers appeared consistent with a purely phonetic effect (this should be the case, given that the Italian stimuli were a subset of the main stimuli, and should be representative of the original findings). Based on the ratio, there appears to be more than simply a phonetic effect with the multisyllabic English. However, no standard limiting value is known to distinguish between what is phonetic and what indicates the
additional presence of a phonological component. At best, the phonological effect seen here for the CVE-English can be noted as minimal.

For the U.S.-born speakers, the CVE magnitude differed significantly as a function of language of utterance. Surprisingly, the magnitude was greater for the Italian than for the English. The size of the CVE-Italian for the U.S.-born speakers, equal to a derived ratio of 1.35, suggests the presence of both a phonetic and a partially suppressed phonological mechanism (as before, this corresponds with the conclusions of the core cases; the Italian stimuli were a subset of the main stimuli set, and were expected to be representative of the prior findings). When English was the language of utterance, the CVE for the U.S.-born speakers was smaller, equal only to a derived ratio of 1.23, suggesting a phonetic effect in combination with a minimal phonological component. The magnitude of the phonological component of the effect may be seen as a small phonological contribution (as compared with the core cases) or as a suppression (as seen with the Italian) of the usual effect size. The ratios for each language, for both speaker groups, are summarized in Table 3.6.

### Table 3.6

<table>
<thead>
<tr>
<th></th>
<th>Foreign-born</th>
<th>U.S.-born</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italian ratio</td>
<td>1.09</td>
<td>1.35</td>
</tr>
<tr>
<td>English ratio</td>
<td>1.16</td>
<td>1.23</td>
</tr>
</tbody>
</table>
3.3 Comparing the core and heterosyllabic cases

The data patterns that have emerged in Sections 3.1 and 3.2 are quite different. It is worth stepping back to compare the results for the core cases with those for the multisyllabic cases and to address the potentially conflicting interpretations of the data patterns.

The ratio for foreign-born speakers for the full set of Italian word pairs was 1.10 (1.09 for the subset), which was characterized as a purely phonetic effect. For the English stimuli, this group exhibited the sizable ratio of 1.45 for the core English cases, which reduced considerably, to 1.16, for the multisyllabic cases. Foreign-born speakers displayed similarly sized absolute duration differences when the stimuli were matched for syllable position of the trigger consonant (16.8 ms [Italian] as compared with 18.8 ms [English]) but, because overall durations were greater for Italian as language of utterance than for English, these duration differences resulted in ratios that could potentially have separate interpretations. While an additional, phonological effect was present with the core cases in English, what may exist beyond the phonetic effect for the multisyllabic English cases remains open to interpretation.

The ratio for U.S.-born speakers for the full set of Italian stimuli was 1.38 (1.35 for the subset). This was interpreted as a phonetic effect in combination with a suppressed phonological effect. With English as the language of utterance, the U.S.-born speakers displayed a large CVE magnitude for the core cases, but then a relatively small magnitude for the multisyllabic cases. The contrast across the two syllabic environments was highly robust: the ratio reduced from 1.66 when the trigger consonant was tautosyllabic to 1.23 when heterosyllabic. In each context, both the phonetic and phonological mechanisms appeared to play a role, albeit a smaller role in the latter, heterosyllabic, context.
Taking the full set of results into consideration, various questions arise that bear on the magnitude of the CVE as found with each language of utterance, for each stimulus set.

- Why did the phonological mechanism not play a (larger) role with the foreign-born speakers in their production of the multisyllabic English stimuli? Or, conversely, why did the phonological mechanism play such a large role with the monosyllabic English stimuli?

- Why did the phonological mechanism not play a larger role with the U.S.-born speakers in their production of the multisyllabic English stimuli?

- Why did the U.S.-born speakers exhibit a greater CVE magnitude with the Italian utterances than with the multisyllabic English?

The questions regarding the phonological mechanism bear on the expectation that the magnified CVE found with English utterances be present for monosyllabic and multisyllabic words alike. There is also a question as to why the CVE would be expected to be greater for English than for Italian, even when the stimuli are matched for syllable position of the trigger consonant. Some of the questions that get asked relate to how the experiment was conducted and the order of presentation of the data. If discussion had instead begun with the heterosyllabic environment, the questions may have been focused on the increased, rather than decreased, magnitude of the CVE when moving from multisyllabic to monosyllabic stimuli (particularly for the foreign-born speakers who exhibit similar patterns in the two languages, in the heterosyllabic environment). Additionally, given the construction of the stimuli, location of the trigger consonant in tautosyllabic versus heterosyllabic position in English was conflated with monosyllabic versus multisyllabic words.
3.3.1 Syllabic effects

The magnitude of the CVE-English with the multisyllabic words, for U.S.-born speakers in particular, may seem surprisingly small given the expected phonological enhancement. However, past research has shown a similar reduction in CVE magnitude for English when a syllable is added to a monosyllabic word, all else being equal (Klatt 1973, Tauberer 2010).

Taking the durational value for a monosyllabic word with a voiced coda consonant (e.g., *need*) as the default “inherent” duration for a given vowel, Klatt (1973) offered a formula to detail the impact on a stress-bearing vowel of adding a syllable to create a disyllabic word (e.g., *needless*) or changing the [voice] value of the trigger consonant to create a voiceless environment (e.g., *neat*). Klatt estimated that either process in isolation would result in a vowel duration equivalent to approximately 66% of the default duration. The separate effect of these processes, applied together, would be expected to result in a duration value on the order of 44% of the default duration if they were independent.

According to Klatt, however, there is a limit to the amount of compression that a stressed vowel can undergo,\(^2\) whether that compression is voicing-related, i.e., a consequence of changing the [voice] value of the trigger consonant, or syllabicity-related, i.e., a consequence of changing the word’s syllabicity from monosyllabic to multisyllabic. Thus, per Klatt’s estimate for his own data, vowel duration is actualized at approximately 54% of the default duration when both processes take place to create a disyllabic word with a voiceless trigger consonant (e.g., *neatly*). The values found with the U.S.-born speakers in the present study bear a striking resemblance to the findings of Klatt, in terms of both absolute durations and relative percentages,

\(^2\) While Klatt referred to each process as a shortening of the intrinsic vowel duration, it may be more appropriate to think of these processes as lengthening events. The motivation for categorizing each process as a lengthening or a shortening is beyond the scope of the present discussion.
as detailed in Table 3.7, below. Values on the left-hand side of the table come from U.S.-born
speakers in the current study; values on the right-hand side come from Klatt (1973).

Table 3.7  Vowel duration means in milliseconds (and percent, with the voiced monosyllable
serving as the default duration) for monosyllabic and multisyllabic words in
English, as a function of [voice] value in the post-vocalic environment.

<table>
<thead>
<tr>
<th></th>
<th>Current study</th>
<th>Klatt (1973)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 syllable</td>
<td>2+ syllables</td>
</tr>
<tr>
<td>voiceless</td>
<td>135 (61%)</td>
<td>116 (52%)</td>
</tr>
<tr>
<td>voiced</td>
<td>222 (100%)</td>
<td>142 (64%)</td>
</tr>
</tbody>
</table>

One interpretation of the results is to propose that differences in syllabicity could explain
the inconsistency in data patterns for the core versus heterosyllabic cases. However, the
variability in the data across speakers, with a range of CVE magnitudes, suggests that syllabicity
differences are not the only explanation. If syllabicity effects were the sole source of the
duration differences across syllable structure, the expectation would be to find similar CVE
magnitudes and ratios — regardless of language of utterance and speaker group — for all the
multisyllabic stimuli. Notably, the foreign-born speakers exhibited similar behavior across the
two languages for the heterosyllabic environment. However, the same pattern did not hold for
the U.S.-born speakers. Instead of a small phonetic effect for the Italian stimulus set, similar to
the 1.10 ratio found with the foreign-born speakers, the ratio for U.S.-born speakers was 1.38.

3.3.2 More general timing patterns

Timing patterns across the two languages may provide an additional clue as to why such
a great amount of compression (or, alternatively, expansion) is found with English as the
language of utterance, but not Italian. Languages are often categorized into rhythmic classes (Roach 1982). These labels are used to capture the differing bases for isochrony in the languages. Generally speaking, Italian tends to be a syllable-timed language (with approximate isochrony from syllable to syllable) and English tends to be a stress-timed language (with approximate isochrony from stress to stress). Syllable-timed languages tend to exhibit less vowel reduction than stress-timed languages. In stress-timed languages, stress tends to manifest itself more prominently than it does in syllable-timed languages (Dauer 1983), with a relevant — but not exclusive — acoustic correlate in English being greater duration. Therefore, more voicing-related compression would be expected with English as the language of utterance, as there is greater freedom to enhance stress with durational variation in stress-based languages.

A question to ask is whether it is possible that the foreign-born speakers are using the same strategy in both Italian and English when the stimuli are matched for syllable structure. Or is the slightly different behavior in the two languages — not statistically, but on the basis of the numerically different derived ratios — indicative of separate processes at work? The foreign-born speakers, due to late acquisition of English, likely have difficulty incorporating the timing patterns consistent with English. As discussed earlier, many of the English-language production errors by the foreign-born speakers involved attempts at re-syllabification of monosyllabic words into disyllabic words, i.e., imposition of a different rhythmic pattern onto the English language stimuli (see §2.2.5). Incorporation of a syllable-based timing approach when uttering the multisyllabic English stimuli, as compared with the stress-based timing used by the U.S.-born

---

25 The idea of a language being grouped into a category such as stress-timed or syllable-timed is a useful but sometimes controversial one. This sort of classification is especially practical for indicating characteristics that are generally found with one timing pattern versus another, e.g., permissible syllable structures. These categorizations are often not considered to be exclusive but rather are tendencies toward a particular pattern.

26 Other characteristics of stress in English include vowel quality, greater loudness and greater pitch excursion (Fry 1958).
speakers, would mean less syllabicity-related compression of the target vowel by the foreign-born speakers than by the U.S.-born speakers. Less syllabicity-related compression permits a larger interval for the two [voice] contexts to emerge, consistent with the findings. It could be proposed that the slightly greater ratio found for the multisyllabic English stimuli is indicative of (partial) incorporation of the phonological effect in combination with syllable-based timing, while the lesser ratio for the Italian stimuli is indicative of an absence of the phonological component.

Having been exposed to Italian since birth, it can be postulated that the U.S.-born speakers have some level of ability (although not perfect ability) to incorporate the timing patterns consistent with their non-dominant language, Italian. With the target vowel-trigger consonant sequence matched for syllabicity across language, the greater syllabicity-related compression found for the target vowel within English utterances meant a relatively smaller CVE magnitude — due to less opportunity for a larger effect following syllabicity-related compression — while the lesser syllabicity-related compression for Italian meant a relatively greater CVE magnitude within Italian utterances. The CVE magnitudes for the U.S.-born speakers appear to reflect the presence of the phonologically enhanced effect as found with English in conjunction with the rhythmic patterns as found with Italian. This topic of expected vowel duration leads to a question regarding the status of the underlying phonological contribution of the trigger consonant, as discussed in the following section.

3.4 Consonantal voicing effect: Flapping

A potential test of the phonological component believed to be present in the underlying representation of English is a comparison of flap and non-flap cases: multisyllabic items that
differ in place of articulation of the trigger consonant. With non-coronal cases, there exists both a phonetic and phonological source for the CVE. Coronal cases are complex in that the triggering consonant is positioned to undergo the English-specific process of “flapping,” resulting in surface \([r]\) from underlying \(/t/\) or \(/d/\). When flapping occurs, the phonetic source of the CVE is neutralized, as the crucial vowel-consonant sequence in both words becomes string identical. While there is some discussion in the literature of the exact nature of the flap, and possible differences in quality based upon whether the underlying segment is \(/t/\) or \(/d/\), the consensus is that the flap is a voiced segment, and is more similar to \([d]\) than to \([t]\). The flapping process leaves just a phonological source of the CVE for coronal place of articulation. Therefore, a smaller CVE magnitude would be expected for coronal, as compared with non-coronal, place of articulation. Analyses with the multisyllabic English data thus far for the present study have only included non-coronal cases.

In comparing the data, foreign-born speakers were excluded from this analysis. These speakers frequently failed to flap, flapping \(/t/\) in less than half of the cases. By comparison, the U.S.-born speakers flapped \(/t/\) 69% of the time. While 3 of the U.S.-born speakers flapped 100% of the time, 2 only flapped 17% of the time. The final speaker for this group flapped 83% of the time. See Appendix H for a list of counts of segments with underlying \(/t/\) categorized by surface realization with corresponding percentages for each participant. The following analysis, therefore, compares vowel duration based on place of articulation of the trigger consonant — coronal or non-coronal — for U.S.-born speakers only.

Since not all speakers flapped all of the time, all cases in which a speaker failed to flap the underlying \(/t/\) were removed, creating a “restricted” data set. Utterances had been transcribed during the segmentation process as described in §2.2.4. The difference between \([t]\) and \([r]\) was
identifiable via the audio playback and was additionally confirmed by spectrographic evidence of a very short closure and a small presence of consonantal phonation for the flaps as compared with the [t]. The difference between [d] and [ɾ] was more difficult to distinguish. Instances in which a speaker failed to flap the underlying /d/ were kept with the data since the [voice] value was consistent with [ɾ].

Vowel duration values for the multisyllabic English stimuli were analyzed for U.S.-born speakers using a two-by-two design having factors Place of Articulation and [Voice] Value. Place of articulation refers to the 6 coronal (e.g., metal, medal) and 6 non-coronal (e.g., soccer, soggy) minimal pairs for the multisyllabic English stimulus set. The non-coronal data are identical to the multisyllabic English values presented in §3.2.

Vowel duration patterns for this data set are summarized in Figure 3.5, below.

![Figure 3.5](image)

**Figure 3.5** Mean duration (and standard error) of critical vowel for the multisyllabic English-language stimuli for U.S.-born speakers, as a function of place of articulation and [voice] value in the post-vocalic environment.

The interaction terms of the ANOVA suggested that the CVE differed in magnitude as a function of place of articulation, $F_1(1,5) = 7.63, p<.05$, $F_2(1,10) = 19.23, p<.01$, but was only reliable with
non-coronal place of articulation: for coronal, $F_1<1$, $F_2(1,5) = 1.68$, $p>.25$, and for non-coronal, $F_1(1,5) = 44.71$, $p<.01$; $F_2(1,5) = 18.44$, $p<.01$.

For coronal place of articulation the CVE was -4.3 ms. Contrary to expectations, durations were numerically greater in the voiceless context, with a ratio of 0.96. With non-coronal place of articulation, the CVE was 26 ms, equal to a ratio of 1.23. This finding was fairly similar to that of Tauberer (2010) who reported a mean ratio of 1.29 (25 ms) for disyllabic words with non-coronal place of articulation. Table 3.8 summarizes the CVEs for coronal and non-coronal place of articulation.

<table>
<thead>
<tr>
<th>Coronal</th>
<th>-4.3 (3.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-coronal</td>
<td>26.0 (6.1)</td>
</tr>
</tbody>
</table>

The question being asked is whether or not there is an underlying effect of [voice] value that survives surface neutralization. The notion of a lengthening occurring despite surface neutralization of the context required to activate the process was addressed in Kenstowicz (1994) as part of a discussion on levels of representation. He argued that the process lengthening vowels preceding a voiced segment must precede the process whereby /t/ and /d/ are flapped. Kenstowicz claimed that more than one level of representation was necessary to account for the lengthening effect found with underlyingly voiced trigger consonants, as these segments were neutralized on the surface by the flapping process. An earlier study by Fox and Terbeek (1977) had a similar take. Interested in determining rule ordering for the processes of vowel lengthening and flapping, they looked at heterosyllabic cases with coronal place of articulation and found vowels to be significantly longer when the [voice] value for the trigger consonant was
underlyingly voiced. They concluded that the CVE was triggered by the underlying [voice] value and, therefore, the vowel lengthening process must precede the flapping process.

The findings for coronal place of articulation did not match the findings of previous studies. Herd, Jongman and Sereno (2010) found a 1.05 ratio (6 ms difference) for disyllabic words with coronal place of articulation, while Tauberer (2010) found a ratio of 1.07 (9.6 ms). Tauberer, however, acknowledged the possibility that not all participants in his study flapped, and since he did not remove the failure-to-flap cases, the effect he found may have been inflated. An analysis of the full data set, prior to removal of instances in which speakers failed to flap the underlying /t/, found the CVE for the full data set to be 8 ms, equivalent to a ratio of 1.05. The CVE for the coronal data, however, was not reliable, $F_1(1,5) = 1.98, p>.10, F_2(1,5) = 4.80, p=.080$.

Comparison of the non-coronal and coronal heterosyllabic stimuli raises the possibility that [voice] in phonological representation may track with pronunciation, and not with the orthography. A fuller investigation into the nature of the underlying representation of [voice] is beyond the scope of this work, but considerations of (speaker and item) power, ordering and lexical representation should all be taken into account. Ideas for exploring the underlying representation will be discussed further in the conclusions chapter.

3.5 Other regions

Measurements were taken for seven regions throughout the utterance, as detailed in §2.2.4. It seems relevant in terms of the data patterns seen in the preceding sections to also include the data related to the closure for the post-vocalic (trigger) consonant. Closure duration is known to typically have an inverse relationship with vowel duration (Lisker 1957, Chen 1970,
Fowler 1992) in that the closure is greater for voiceless consonants and lesser for voiced consonants.

The investigator chose not to present the values for the prevocalic consonant due to the large variability in phonetic content of the segment, e.g., Italian *prete, tubo*; English: *cup, frog*. While the materials design dictated uniformity of post-vocalic consonants, prevocalic consonants were less constrained, other than the requirement to be consistent with the other member of the minimal pair. The post-vocalic consonant was also set aside for purposes of the current analysis. Both regions are suitable for future study.

### 3.5.1 Closure: Core cases

Closure duration values for the post-vocalic consonant were analyzed for the core cases using a three-way design having factors Speaker Group, Language of Utterance and [Voice] Value. Language of utterance refers to the 18 minimal pairs for the Italian stimulus set and the 18 minimal pairs for the English stimulus set as described in §3.1.

Closure duration patterns for both speaker groups are summarized in Figure 3.6, below.
Figure 3.6  Mean duration (and standard error) of post-vocalic consonant closure for foreign- and U.S.-born speaker groups, as a function of language of utterance and [voice] value in the post-vocalic environment.

There was a three-way interaction of speaker group, language of utterance and [voice] value, $F_1(1,11) = 14.53, p<.01, F_2(1,34) = 35.79, p<.001$. As the figure suggests, data patterns for the two speaker groups differ as a function of language of utterance and [voice] value.

To further explore the data patterns in the differing [voice] contexts, two-way subanalyses were conducted for each speaker group. Data for the foreign-born participants are presented in the left panel of Figure 3.6. Closure duration was found to be significantly greater when the trigger consonant was voiceless rather than voiced, $F_1(1,6) = 42.89, p<.001, F_2(1,34) = 112.07, p<.001$. Closure duration was also greater for this group when English was the language of utterance as compared with Italian, $F_1(1,6) = 9.82, p<.05, F_2(1,34) = 27.57, p<.001$.

Data for the U.S.-born participants are presented in the right panel of Figure 3.6. With this speaker group, subanalyses revealed a two-way interaction of language of utterance and [voice] value, $F_1(1,5) = 195.97, p<.001, F_2(1,34) = 31.95, p<.001$, suggesting that the difference
in closure durations across [voice] value differed in magnitude for the U.S.-born speakers as a function of language of utterance. Further subanalyses confirmed the difference to be reliable both for the Italian, $F_1(1,5) = 55.29$, $p<.001$, $F_2(1,17) = 156.72$, $p<.001$, and for the English, $F_1(1,5) = 9.58$, $p<.05$, $F_2(1,17) = 58.39$, $p<.001$; closure durations were greater in the voiceless context.

The data patterns seen with these data suggested that the inverse vowel-closure relationship was present for both groups of bilingual speakers in each language of utterance. The magnitude of the difference appears to be greatest for the U.S.-born speakers with Italian utterances. We look now at the closure values when the stimuli were matched for syllable structure.

### 3.5.2 Closure: The heterosyllabic environment

Closure duration values were analyzed for the multisyllabic stimuli using a three-way design having factors Speaker Group, Language of Utterance and [Voice] Value. With this analysis, language of utterance refers to the subset of 6 minimal pairs for the Italian stimulus set and the 6 non-coronal minimal pairs for the English stimulus set as described in §3.2.

Closure duration patterns for both speaker groups are summarized below in Figure 3.7.
The overall analysis of variance revealed a three-way interaction of speaker group, language of utterance and [voice] value, $F_1(1,11) = 21.71$, $p<.01$, $F_2(1,10) = 10.1$, $p<.01$.

Two-way subanalyses were conducted within each speaker group to fully examine the data patterns. We first look at the data for the foreign-born speakers, presented in the left panel of Figure 3.7. Closure duration was found to be significantly greater when the trigger consonant was voiceless rather than voiced, $F_1(1,6) = 32.93$, $p<.01$, $F_2(1,10) = 27.14$, $p<.001$. No difference was found when the language of utterance was Italian versus English, $Fs < 1$ in both subject- and item-based analyses.

Next we look at the data for the U.S.-born participants, presented in Figure 3.7’s right panel. The interaction terms of the ANOVA showed that the duration difference corresponding to [voice] value of the trigger consonant differed in magnitude as a function of the language of utterance, $F_1(1,5) = 80.28$, $p<.001$, $F_2(1,10) = 11.23$, $p<.01$. The difference was reliable for both
the Italian, $F_1(1,5) = 174.05$, $p<.001$, $F_2(1,5) = 48.07$, $p<.001$, and the English, $F_1(1,5) = 18.83$, $p<.01$, $F_2(1,5) = 89.77$, $p<.001$, with closure durations being greater when the [voice] value was voiceless, as compared with voiced.

The data patterns seen in Figure 3.7 suggested that, as with the core cases, the inverse vowel-closure relationship was present for both groups of bilingual speakers in each language of utterance. And again, the magnitude of the difference appears to be greatest for the U.S.-born speakers with Italian utterances. The Italian stimuli were a subset of the full stimulus set; hence, the closure data patterns for the Italian were similar here to those found with the core cases.

The patterns with the Italian data appear most interesting: just as the U.S.-born speakers differentiate vowel duration as a function of [voice] value to a greater extent than do the foreign-born speakers, they also differentiate more with the closure. This might be related to the rhythmic timing specific to Italian. Farnetani and Kori (1986) concluded that the timing in Italian is organized around a “rhythmical syllable” which extends from vowel onset to vowel onset, and that these timing units tend to be fairly isochronous. It might be that the U.S.-born speakers maintain rhythmic timing patterns in Italian such that the voiceless trigger consonant corresponds with a somewhat exaggerated closure duration while the voiced trigger consonant corresponds with a relatively smaller closure duration. The greater or smaller duration seen for the closure region with the U.S.-born speakers would then, in essence, compensate for the respectively smaller or greater duration of the target vowel to achieve relative parity in duration from vowel onset to vowel onset.
3.6 Summary of findings

The findings presented throughout this chapter illustrate how two groups of bilingual speakers with varying degrees of fluency in each of their languages utilize the consonantal voicing effect. Data patterns for the core cases painted a picture in which the CVE was always of greater magnitude in terms of both the absolute duration and the ratio with English as the language of utterance, cf. Italian as the language of utterance, and always greater for the U.S.-born speakers, cf. the foreign-born speakers. The foreign-born speakers maintained a solely phonetic effect with the Italian-language stimuli, but managed an enhancement, consistent with a phonological component, for the English-language stimuli. The U.S.-born speakers displayed a CVE magnitude consistent with a combination of a phonetic effect and a greatly enhanced phonological component when English, their dominant language, was the language of utterance. When Italian was the language of utterance, the U.S.-born speakers managed to partially suppress the phonological enhancement.

Following the introduction of the multisyllabic stimuli, where the stimuli sets were matched for trigger consonant syllable position, a deeper level of interpretation of the results became necessary. As with the core cases, the CVE was present for both groups of bilingual speakers in each language of utterance. With the foreign-born speakers, however, the CVEs were similar across the two languages of utterance, corresponding to a decrease in magnitude of the CVE-English for the core cases. The CVE-English was also smaller for the U.S.-born speakers with these data. Further, the CVE-English for the U.S.-born speakers was significantly smaller than the CVE-Italian.

The difference in data patterns appears to be about both the languages and the speaker groups. To account for the patterns seen with each stimulus set, the discourse necessarily
incorporates timing patterns. It is also necessary to understand the extent to which elements of one’s dominant language may influence a speaker’s production in the non-dominant language. The following chapter will discuss the implications of these findings.
Chapter 4: Concluding remarks

The goal of this study is to quantify the consonantal voicing effect (CVE) for two groups of bilingual speakers of Italian and English who differed in their language experience. With the first group, foreign-born speakers of Italian with late-acquired English, Italian was the main language from birth until the time of their emigration from Italy, in their teenage or early adult years. With the U.S.-born group, speakers had early exposure to both Italian and English but once they reached school age, English was always the dominant language in that it was the language of education, the language of majority-use in daily life and, according to self-ratings, the language of greater proficiency. We return in this chapter to our original question of how CVE settings are maintained by bilingual speakers as a function of their experience with the two languages.

4.1 Consonantal voicing effects in dominant and non-dominant languages

A main finding with this study is that speakers were able to maintain native-like timing settings for the CVE with their dominant language, which was Italian in the case of the foreign-born speakers and English in the case of the U.S.-born speakers. While timing settings in the non-dominant language differed from those found in the dominant language for both speaker groups, native-like settings in the non-dominant language were not attained by either group.

The contrast in the two [voice] contexts is much more robust for English than it is for Italian. The timing settings found in this study for the bilingual’s two languages were strongly impacted in all cases by the dominant language. Language dominance appears to contribute to maintenance of native-like settings in an L1 even when an L2 is in frequent usage.
in the L2 show clear influence of the L1: the mid-sized effect for the L2 as observed for the core cases was relatively similar for both groups.

### 4.1.1 Core data: Italian multisyllabic words, English monosyllabic words

The patterns seen with the core sets of data — those derived from Italian multisyllabic and English monosyllabic words, i.e., the word structures most typical of each of those languages — show that the CVE was present for both groups of bilingual speakers in each language. But, crucially, the magnitude of the CVE differed by group and by language: the CVE was always greater when English was the language of utterance as compared with Italian; and the CVE was always greater with the U.S.-born speakers as compared with the foreign-born speakers. The size of the presumed phonological enhancement found in the CVE in each instance is of greatest interest.

Foreign-born speakers, as anticipated, maintained a native-like setting for the CVE in Italian, their first language: that is, the magnitude of the CVE was consistent with values that have been reported in the literature. All participants from this group communicate regularly in Italian, which is self-rated to be their dominant language. While two of the speakers expressed greater “comfort” using English than Italian, all speakers rated themselves as having better overall skills in Italian versus English in the areas of listening, speaking, reading and writing. Despite their many years in the U.S., there did not appear to be any interference from the second language in that there was no carryover from their L2 English into their L1 Italian.

The foreign-born speakers had had a long period of immersion in an English-speaking environment, the impact of which appears in their settings for the core English-language cases. Their CVE-English approached, but did not reach, the range that is considered the norm for the
default cases in English: that is, the foreign-born speakers were undershooting in English, not fully reaching the target timing pattern. This group managed to implement to some degree the phonological enhancement that is presumed to be present for English, but not to the extent previously reported for monolingual speakers. The ability to reproduce the enhanced effect but not to its full extent is likely attributable to the age of the speakers on arrival in the U.S., when they began using English in combination with their continued usage of Italian.

The U.S.-born speakers were found to have a native-like setting for the CVE in their dominant language, English: that is, the magnitude of the CVE was entirely compatible with previously established norms. As with the foreign-born group, there did not appear to be any interference in the dominant language from the L2. In their non-dominant language, Italian, U.S.-born speakers exhibited a setting for the CVE that fell between the established language-specific settings. It appears that when speaking Italian, this group was unable to fully suppress the phonological enhancement found for the CVE in English, meaning that there was a carryover effect from the dominant language. The CVE was partially reduced due to the syllable structure for the Italian stimuli, which was multisyllabic as compared with the monosyllabic structure for the core English cases. But it appears that maintenance of the timing patterns consistent with Italian as a syllable-timed language meant a lesser degree of CVE-related compression than would be found with English, a stress-based language.

4.1.2 Additional data: Multisyllabic words

Once the stimuli sets for each language were matched for the trigger consonant’s syllable position, the data patterns showed that, consistent with the core cases, the CVE was present for both groups of bilingual speakers in each language of utterance. The area of interest for
discussion with these cases was settings in each speaker group for the English-language stimuli, where for the first time stimulus items were no longer monosyllabic; note that multisyllabicity was necessarily a constant for Italian.

For the foreign-born speakers, the CVE-English for multisyllabic stimuli was greatly reduced in magnitude from the value that had been found for monosyllabic stimuli. Duration differences across the two languages were small, and did not differ statistically. But derived ratios differed enough across the two languages that a question arose as to whether something different was happening in the two languages. The numerically larger ratio with the English may reflect the presence — albeit small — of a phonological enhancement.

The CVE-English found for the U.S.-born speakers with the multisyllabic stimuli was also reduced from that found for monosyllabic stimuli. With such cases, the Italian-language ratio was greater than the English-language ratio. As noted previously, the U.S.-born speakers were unable to suppress entirely the phonological enhancement with the Italian-language stimuli. The CVE-Italian appears to reflect an interplay of the phonologically enhanced effect as found with English in combination with the rhythmic patterns as found with Italian. The smaller CVE (and smaller ratio) found for the U.S.-born speakers with English as compared with Italian appears to result from the greater compression found in languages with stress-based timing as opposed to syllable-based timing.

An issue that arises is that, given the construction of the stimuli, location of the trigger consonant in tautosyllabic versus heterosyllabic position in English was conflated with monosyllabic versus multisyllabic words. This leads us to ask whether differences in estimated CVE values are solely attributable to syllabicity or whether, alternatively, the differences should be attributed to the location of the trigger consonant in the coda of the same syllable (in words
that are monosyllables) versus the onset of the following syllable (in words of greater syllabicity). We take the issue to be settled by Tauberer’s (2010) direct evaluation of the trigger consonant’s location. In his study of multisyllabic English-language stimuli, Tauberer reported no systematic difference in the magnitude of the CVE across tautosyllabic and heterosyllabic environments, e.g., fraction [frækʃən] – fragment [fræg.mənt], cf. backing [bækɪŋ] – bagging [bægɪŋ].

4.1.3 Data for the case of flapping

The effect of the CVE with coronal cases, which are subject to flapping, was expected to differ from the effect with non-coronal cases, which are not subject to flapping. This was consistent with our findings. Crucially, there is no reason to believe that the CVE would be intrinsically smaller for coronal cases as compared with non-coronal cases when the flapping environment is not present (see, e.g., House & Fairbanks 1953). Once a segment is flapped, however, the context for the CVE to occur has been neutralized. It is then assumed that any phonetic variation has been removed, but that the underlying phonological element remains. In the present study, once the presumed underlying distinction in [voice] value was surface-neutralized via flapping, any residual, purely phonological effect was not statistically detectable. This finding was not consistent with prior studies, most of which show a modest contribution of the CVE.

We offer two possible explanations for the lack of an observed effect in these data, reverse ordering and lexicalization. On the one hand, it may be that the rule-ordering that has been traditionally assumed for these processes (where vowel lengthening precedes flapping, as per Kenstowicz’s opacity argument), does not match the ordering actualized by this population.
The results found for this study would instead match an ordering in which the flapping process precedes the vowel lengthening process. On the other hand, it may be that lexicalization is driving the lack of duration differences. As suggested in §3.4, a more comprehensive examination of flapping could give a better understanding of the underlying representation of [voice] as it interacts with the CVE. That more comprehensive study could incorporate words formed by highly productive affixation, i.e., inflection as in writing ~ riding, which are less likely to be lexicalized with flap in the underlying form than are the products of derivational affixation, e.g., writer ~ rider. Incorporating a context in which the base form, e.g., write, is presented prior to utterance of the target word would ensure that the underlying [voice] value would be in place. Further, increased power (more speakers, more exemplars) would increase the likelihood that even a rather small CVE could be detected.

4.2 Alternative approaches with the current study

Complementary future work could incorporate an additional speaker group of Italian-English bilinguals. As reported by Zampini and Green (2001) in their study of VOT settings, no significant difference was found for early-bilingual speakers of Spanish and English from those of monolingual speakers of the two languages. The Spanish-English speakers had frequent (daily) exposure to both languages. Inclusion of early-bilinguals with more consistent exposure to the non-dominant language following childhood (perhaps via travel, study-abroad and/or work experiences) could provide further insight into the efficacy of the CVE as a predictor of native-like skills. It may turn out that certain phonetic settings are more malleable than others as has been suggested by some researchers (Wilson & Gick 2014, among others) and that VOT is a
more malleable setting than vowel duration, or that age of learning would continue to the best predictor of the differences found for the CVE. One of the potential issues with respect to identifying early bilinguals is the difficulty in locating participants with a common region of origin to control for possible regional variation in speech.

Another option, particularly if adding a more balanced group of bilinguals than was utilized with the current study, would be to measure perceived accent in the two languages based on judgments by native speakers, similar to that conducted by Wilson and Gick (2014). This could make clear whether a native-like setting for the CVE is necessary to sound accent-free. This work consequently has applications in the area of language pedagogy, particularly in the area of pronunciation, as language-specific CVE settings may contribute to a speaker’s intelligibility in a second language.

A final area of interest is the basis for timing in the second language. A fuller investigation of the cross-linguistic timing differences in terms of syllable- versus stress-timed languages using the current set of data could provide insight into how speakers manage the L2. The data could be analyzed to better understand the relationship between syllable structure and timing patterns, and the resulting impact on the CVE.

4.3 Directions for future research

As mentioned previously, an issue identified is that there is a conflation of two stimulus properties, one being syllabic (monosyllabic versus multisyllabic) and the other being position of the trigger consonant (tautosyllabic versus heterosyllabic). The effect of syllable structure for these two groups of bilingual speakers can be explored by setting aside the voicing contrast while comparing CV.CV to CVC.CV. For the English, this means looking at vowel duration
when the target vowel is in an open syllable compared with a closed syllable, e.g., *chapel* ~ *chapter*, addressing the question of durational consequences of a heterosyllabic versus a tautosyllabic trigger consonant in the multisyllabic environment. For the Italian, the equivalent contrast is achieved by examining vowel durations in singleton versus geminate contexts, e.g., *fato* [fa.to] ‘fate’ ~ *fatto* [fat.to] ‘fact; made.’

Another question that may be worth addressing is the role of [voice] value of the prevocalic consonant. The position taken by Peterson & Lehiste (1960) was that the [voice] value of the prevocalic consonant does not have a detectable impact on vowel duration, but Hillenbrand et al. (2001) found otherwise; see §1.1.1 for further discussion. The impact of [voice] value of the prevocalic consonant on vowel duration appears to be an under-examined area, cross-linguistically, likely because the effect has been reported to be minimal at best, and has little perceptual consequence, i.e., it does not appear to be a cue for voicing quality in the prevocalic consonant in the same way as has been reported for the post-vocalic consonant. Nonetheless, the impact of the prevocalic consonant for both Italian and English, assessed by comparing minimal pairs with contrastive voicing in the consonant preceding the target vowel, e.g, English *patch* ~ *batch*; Italian *pollo* ‘chicken’ ~ *bollo* ‘stamp,’ may reveal previously unreported cross-linguistic differences, and at the very least would help to complete the picture of elements contributing to timing patterns in these languages.27

The CVE has been used successfully here to characterize bilingualism through phonetic variation. What we have seen is that language dominance plays a large role in the patterns found, as the patterns are often influenced by the stronger language. The general trend in

---

27 Utterances exemplifying both the stimulus contrasts described above — examining the effect of syllable structure, and the impact of the prevocalic consonant’s [voice] value — were in fact recorded in the course of the current study; such cases fell among the experiment’s fillers. However, data extraction and analysis was beyond the scope of the current dissertation.
research has been that, once a second language is at issue, it becomes the target of primary interest. The first language of bilingual speakers often gets neglected, other than as a basis of comparison for production in the second language. We hope that this study, along with other recent studies that look at both — or all — of a speaker’s languages, helps reverse that trend.
Appendices

Appendix A  Figure from Flege, Munro and MacKay (1995): Vowel duration and stop closure duration

The figure below, reproduced directly from Flege, Munro and MacKay (1995), provides values for native speakers of Canadian English (AOA=0) and for native speakers of Italian as a function of age of arrival (AOA) in the U.S. Values are given in the top panel for vowel duration and in the bottom panel for closure duration, for two English language word pairs, *tag ~ tack* and *tagging ~ tacking*.

For further detail, refer to Section 1.4.3 in Chapter 1.
Appendix B  Language History Questionnaire

BACKGROUND INFORMATION

All personal information you will provide is confidential. Feel free to use the back of the last sheet if you need more room.

Age:..........   Sex:   □ male   □ female   Participant No.:   .................

City/Country of origin:   .........................................................   Occupation:   ..............................................................

What language(s) does your mother speak?   ................................       your father?   ......................................................

How old were you when you started to learn English?   ...........................   Italian?   ...........................................................

If both English and Italian were spoken in your home from the time of your birth, check here:   □

Briefly explain when you began learning English and Italian:

English:   ..............................................................................

Italian:   ..............................................................................

Briefly explain where you began learning English and Italian:

English:   ..............................................................................

Italian:   ..............................................................................

Briefly explain with whom you began learning English and Italian:

English:   ..............................................................................

Italian:   ..............................................................................

Do you speak any languages other than English and Italian? (Please say how well you speak any of the languages you list here, e.g. “fluently”, “only a little”).

Have you ever lived outside of the continental United States?

□ No.

□ Yes. If yes, how old were you when you first arrived to the continental US?   ......................................................

Indicate where you lived outside of the continental US, when, and for how long:   ......................................................

Have you spent any time longer than two months living in an environment where English or Italian is not the majority language? □ No.

□ Yes. If yes, how old were you when this happened?   ......................................................

Describe where, when, and for how long:   ......................................................

Education background (check all that apply, and please list the language, if applicable, on the right):

   elementary school   □ in English   □ in Italian   □ in another language

   junior high (middle) school   □ in English   □ in Italian   □ in another language

   high-school   □ in English   □ in Italian   □ in another language

   college   □ in English   □ in Italian   □ in another language

   graduate school   □ in English   □ in Italian   □ in another language

Location (check all that apply, and please list the place, if applicable, on the right):

   Where did you attend elementary school? □ in the continental US □ in Italy □ elsewhere

   Where did you attend junior high (middle) school? □ in the continental US □ in Italy □ elsewhere

   Where did you attend high-school? □ in the continental US □ in Italy □ elsewhere

   Where did you go to college? □ in the continental US □ in Italy □ elsewhere
<table>
<thead>
<tr>
<th>When you were a child, how much Italian/English did you speak:</th>
<th>always Italian</th>
<th>always English</th>
</tr>
</thead>
<tbody>
<tr>
<td>at home, to your parents or primary care-givers?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>at home, to your brothers or sisters?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>at home, to your grandparents?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>at home, to other relatives?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>to your friends?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>in other social contexts (to neighbors, people at the supermarket, etc.)?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>When you were a child, how much Italian/English did the following people speak to you?</th>
<th>always Italian</th>
<th>always English</th>
</tr>
</thead>
<tbody>
<tr>
<td>your parents or primary care-givers</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>your brothers or sisters</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>your grandparents</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>other relatives</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>your friends</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>other people</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>When you were a teenager, how much Italian/English did you speak:</th>
<th>always Italian</th>
<th>always English</th>
</tr>
</thead>
<tbody>
<tr>
<td>at home, to your parents or primary care-givers?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>at home, to your brothers or sisters?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>at home, to your grandparents?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>at home, to other relatives?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>to your friends?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>to your teachers at school?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>in other social contexts?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>When you were a teenager, how much Italian/English did the following people speak to you?</th>
<th>always Italian</th>
<th>always English</th>
</tr>
</thead>
<tbody>
<tr>
<td>your parents</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>your brothers or sisters</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>your grandparents</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>other relatives</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>your friends</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>your teachers at school</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>other people</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How much Italian/English do you speak now, as an adult:</th>
<th>always Italian</th>
<th>always English</th>
</tr>
</thead>
<tbody>
<tr>
<td>at home, to your spouse, living companion, roommate?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>at home, to your children?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>at home, to your younger relatives (siblings, cousins, nieces/nephews, etc.)?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>at home, to your older relatives (parents, grandparents, etc.)?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>to your intimate friends</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>to your colleagues at work/school?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>in other social contexts?</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How much Italian/English do the following people speak to you now, as an adult?</th>
<th>always Italian</th>
<th>always English</th>
</tr>
</thead>
<tbody>
<tr>
<td>your spouse, living companion, roommate</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>your children</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>your younger relatives (siblings, cousins, nieces/nephews, etc.)</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>your older relatives (parents, grandparents, etc.)</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>your intimate friends</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>your colleagues at work/school</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
<tr>
<td>other people</td>
<td>1 2 3 4 5 n/a</td>
<td></td>
</tr>
</tbody>
</table>
Rate yourself according to the following categories (circle one on each line):

How would you rate your overall listening skill in English/Italian?
- English: very good  somewhat good  so-so  somewhat poor  very poor
- Italian: very good  somewhat good  so-so  somewhat poor  very poor

How would you rate your overall speaking ability in English/Italian?
- English: very good  somewhat good  so-so  somewhat poor  very poor
- Italian: very good  somewhat good  so-so  somewhat poor  very poor

How would you rate your overall reading ability in English/Italian?
- English: very good  somewhat good  so-so  somewhat poor  very poor
- Italian: very good  somewhat good  so-so  somewhat poor  very poor

How would you rate your overall writing ability in English/Italian?
- English: very good  somewhat good  so-so  somewhat poor  very poor
- Italian: very good  somewhat good  so-so  somewhat poor  very poor

Overall, which language do you feel more comfortable using?  
☐ English  ☐ Italian

Do you have any other comments on your language background which you think are important but which you were not asked about in this questionnaire?
................................................................................................................................................................................
................................................................................................................................................................................
................................................................................................................................................................................
................................................................................................................................................................................
................................................................................................................................................................................
................................................................................................................................................................................
................................................................................................................................................................................

Thank You for Your Cooperation!
Please take a moment now to ensure that you have filled in all the blanks.
Appendix C  Materials listing: Italian

**Italian, core stimuli**

The table presents word pairs for Italian, categorized by target vowel height and place of articulation of CVE-triggering consonant. The voiceless member of the (near-)minimal pair is in the first column, and the voiced member in the second.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Height</th>
<th>Place</th>
</tr>
</thead>
<tbody>
<tr>
<td>rito</td>
<td>rida</td>
<td>High</td>
<td>Coronal</td>
</tr>
<tr>
<td>lite</td>
<td>lido</td>
<td>High</td>
<td>Coronal</td>
</tr>
<tr>
<td>muta</td>
<td>muda</td>
<td>High</td>
<td>Coronal</td>
</tr>
<tr>
<td>tipo</td>
<td>tubo</td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>lupo</td>
<td>liba</td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>cupola</td>
<td>cubico</td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>fretta</td>
<td>freddo</td>
<td>Mid</td>
<td>Coronal</td>
</tr>
<tr>
<td>moto</td>
<td>modo</td>
<td>Mid</td>
<td>Coronal</td>
</tr>
<tr>
<td>prete</td>
<td>preda</td>
<td>Mid</td>
<td>Coronal</td>
</tr>
<tr>
<td>metodo</td>
<td>medico</td>
<td>Mid</td>
<td>Coronal</td>
</tr>
<tr>
<td>reca</td>
<td>lega</td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>roco</td>
<td>rogo</td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>grata</td>
<td>grado</td>
<td>Low</td>
<td>Coronal</td>
</tr>
<tr>
<td>rata</td>
<td>rado</td>
<td>Low</td>
<td>Coronal</td>
</tr>
<tr>
<td>statua</td>
<td>stadio</td>
<td>Low</td>
<td>Coronal</td>
</tr>
<tr>
<td>sacro</td>
<td>sagra</td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>lapide</td>
<td>labile</td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>brache</td>
<td>drago</td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
</tbody>
</table>

*Note.* Bolded items denote the word pairs comprising the 6-pair subset of the Italian stimulus set. For further detail, refer to Section 3.2.
### Italian, fillers (set 1)

The table presents word pairs for Italian, categorized by target vowel height and place of articulation of post-vocalic consonant. The trigger consonant is in onset position (of the following syllable) for the first member of the pair, and in coda position (of the same syllable, creating a geminate consonant) for the second member of the pair.

<table>
<thead>
<tr>
<th>Height</th>
<th>Place</th>
<th>Word 1</th>
<th>Word 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Coronal</td>
<td>dito</td>
<td>ditta</td>
</tr>
<tr>
<td>High</td>
<td>Coronal</td>
<td>bruto</td>
<td>brutto</td>
</tr>
<tr>
<td>High</td>
<td>Coronal</td>
<td>vita</td>
<td>vitto</td>
</tr>
<tr>
<td>High</td>
<td>Non-Coronal</td>
<td>fico</td>
<td>ficca</td>
</tr>
<tr>
<td>High</td>
<td>Non-Coronal</td>
<td>libro</td>
<td>libbra</td>
</tr>
<tr>
<td>High</td>
<td>Non-Coronal</td>
<td>bibita</td>
<td>bibbia</td>
</tr>
<tr>
<td>Mid</td>
<td>Coronal</td>
<td>nota</td>
<td>notte</td>
</tr>
<tr>
<td>Mid</td>
<td>Coronal</td>
<td>loto</td>
<td>lotta</td>
</tr>
<tr>
<td>Mid</td>
<td>Coronal</td>
<td>greto</td>
<td>gretto</td>
</tr>
<tr>
<td>Mid</td>
<td>Non-Coronal</td>
<td>dopo</td>
<td>doppio</td>
</tr>
<tr>
<td>Mid</td>
<td>Non-Coronal</td>
<td>copia</td>
<td>coppia</td>
</tr>
<tr>
<td>Mid</td>
<td>Non-Coronal</td>
<td>topo</td>
<td>toppa</td>
</tr>
<tr>
<td>Low</td>
<td>Coronal</td>
<td>patria</td>
<td>patto</td>
</tr>
<tr>
<td>Low</td>
<td>Coronal</td>
<td>fato</td>
<td>fatto</td>
</tr>
<tr>
<td>Low</td>
<td>Coronal</td>
<td>lato</td>
<td>latte</td>
</tr>
<tr>
<td>Low</td>
<td>Non-Coronal</td>
<td>capo</td>
<td>cappa</td>
</tr>
<tr>
<td>Low</td>
<td>Non-Coronal</td>
<td>sacro</td>
<td>sacco</td>
</tr>
<tr>
<td>Low</td>
<td>Non-Coronal</td>
<td>papa</td>
<td>pappa</td>
</tr>
<tr>
<td>Mid</td>
<td>Coronal</td>
<td>tono</td>
<td>tonno</td>
</tr>
<tr>
<td>Mid</td>
<td>Coronal</td>
<td>dono</td>
<td>donna</td>
</tr>
<tr>
<td>Mid</td>
<td>Coronal</td>
<td>nona</td>
<td>nonna</td>
</tr>
<tr>
<td>Mid</td>
<td>Coronal</td>
<td>pena</td>
<td>penna</td>
</tr>
<tr>
<td>Mid</td>
<td>Coronal</td>
<td>rene</td>
<td>renna</td>
</tr>
<tr>
<td>Low</td>
<td>Coronal</td>
<td>cane</td>
<td>canne</td>
</tr>
</tbody>
</table>
**Italian, fillers (set 2)**

The table presents word pairs for Italian, categorized by target vowel height and place of articulation of pre-vocalic consonant. The voiceless member of the (near-)minimal pair is in the first column, and the voiced member in the second.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Height</th>
<th>Place</th>
</tr>
</thead>
<tbody>
<tr>
<td>cita</td>
<td>gita</td>
<td>High</td>
<td>Coronal</td>
</tr>
<tr>
<td>finto</td>
<td>vinto</td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>capire</td>
<td>cabina</td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>supino</td>
<td>subire</td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>filo</td>
<td>vile</td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>fìno</td>
<td>vino</td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>tetto</td>
<td>detto</td>
<td>Mid</td>
<td>Coronal</td>
</tr>
<tr>
<td>foto</td>
<td>voto</td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>pollo</td>
<td>bollo</td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>pelle</td>
<td>bella</td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>festa</td>
<td>vespa</td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>pere</td>
<td>bere</td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>cara</td>
<td>gara</td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>passo</td>
<td>basso</td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>palla</td>
<td>balla</td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>fase</td>
<td>vaso</td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>campo</td>
<td>gamba</td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>pace</td>
<td>baci</td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
</tbody>
</table>
Italian, practice items

acrobata
acropoli
basta
carro
oca
ridda
ritto
salute
società
sogno
strada
vicolo
## Appendix D  Materials listing: English

### English, core stimuli

The table presents word pairs for English, categorized by target vowel height and place of articulation of CVE-triggering consonant. The voiceless member of the (near-)minimal pair is in the first column, and the voiced member in the second.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Height</th>
<th>Place</th>
</tr>
</thead>
<tbody>
<tr>
<td>brute</td>
<td>brood</td>
<td>High</td>
<td>Coronal</td>
</tr>
<tr>
<td>root</td>
<td>rude</td>
<td>High</td>
<td>Coronal</td>
</tr>
<tr>
<td>seat</td>
<td>seed</td>
<td>High</td>
<td>Coronal</td>
</tr>
<tr>
<td>wick</td>
<td>wig</td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>pick</td>
<td>pig</td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>rip</td>
<td>rib</td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>great</td>
<td>grade</td>
<td>Mid</td>
<td>Coronal</td>
</tr>
<tr>
<td>debt</td>
<td>dead</td>
<td>Mid</td>
<td>Coronal</td>
</tr>
<tr>
<td>bet</td>
<td>bed</td>
<td>Mid</td>
<td>Coronal</td>
</tr>
<tr>
<td>cup</td>
<td>cub</td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>peck</td>
<td>peg</td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>tuck</td>
<td>tug</td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>not</td>
<td>nod</td>
<td>Low</td>
<td>Coronal</td>
</tr>
<tr>
<td>mat</td>
<td>mad</td>
<td>Low</td>
<td>Coronal</td>
</tr>
<tr>
<td>write</td>
<td>ride</td>
<td>Low</td>
<td>Coronal</td>
</tr>
<tr>
<td>cap</td>
<td>cab</td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>frock</td>
<td>frog</td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>tack</td>
<td>tag</td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
</tbody>
</table>
**English, multisyllabic stimuli**

The table presents word pairs for English, categorized by target vowel height and place of articulation of CVE-triggering consonant. The voiceless member of the (near-)minimal pair is in the first column, and the voiced member in the second.

<table>
<thead>
<tr>
<th>word1</th>
<th>word2</th>
<th>Height</th>
<th>Place</th>
</tr>
</thead>
<tbody>
<tr>
<td>metal</td>
<td>medal</td>
<td>Mid</td>
<td>Coronal</td>
</tr>
<tr>
<td>duty</td>
<td>doodle</td>
<td>High</td>
<td>Coronal</td>
</tr>
<tr>
<td>witty</td>
<td>widow</td>
<td>High</td>
<td>Coronal</td>
</tr>
<tr>
<td>writer</td>
<td>rider</td>
<td>Low</td>
<td>Coronal</td>
</tr>
<tr>
<td>matter</td>
<td>madder</td>
<td>Low</td>
<td>Coronal</td>
</tr>
<tr>
<td>grater</td>
<td>grader</td>
<td>Mid</td>
<td>Coronal</td>
</tr>
<tr>
<td>capital</td>
<td>cabinet</td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>staple</td>
<td>stable</td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>baker</td>
<td>bagel</td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>soccer</td>
<td>soggy</td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>rapid</td>
<td>rabbit</td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td>lucky</td>
<td>luggage</td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
</tbody>
</table>
**English, fillers (set 1)**

The table presents word pairs for English, categorized by target vowel height and place of articulation of post-vocalic consonant. The trigger consonant is in onset position (of the following syllable) for the first member of the pair, and in coda position (of the same syllable) for the second member of the pair.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Height</th>
<th>Place</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>people</strong></td>
<td><strong>peephole</strong></td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>steeple</strong></td>
<td><strong>steepness</strong></td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>nickle</strong></td>
<td><strong>nickname</strong></td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>pickle</strong></td>
<td><strong>picnic</strong></td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>typical</strong></td>
<td><strong>tipsy</strong></td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>cigarette</strong></td>
<td><strong>signature</strong></td>
<td>High</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>second</strong></td>
<td><strong>section</strong></td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>rugged</strong></td>
<td><strong>rugby</strong></td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>couple</strong></td>
<td><strong>cupcake</strong></td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>suburb</strong></td>
<td><strong>subway</strong></td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>wreckage</strong></td>
<td><strong>rectangle</strong></td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>replica</strong></td>
<td><strong>reptile</strong></td>
<td>Mid</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>chapel</strong></td>
<td><strong>chapter</strong></td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>lobby</strong></td>
<td><strong>lobster</strong></td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>khaki</strong></td>
<td><strong>cactus</strong></td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>magazine</strong></td>
<td><strong>magnify</strong></td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>poppy</strong></td>
<td><strong>popcorn</strong></td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
<tr>
<td><strong>tacky</strong></td>
<td><strong>taxi</strong></td>
<td>Low</td>
<td>Non-Coronal</td>
</tr>
</tbody>
</table>
**English, fillers (set 2)**

The table presents word pairs for English, categorized by target vowel height and place of articulation of pre-vocalic consonant. The voiceless member of the (near-)minimal pair is in the first column, and the voiced member in the second.

<table>
<thead>
<tr>
<th>Height</th>
<th>Place</th>
<th>English, fillers (set 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Coronal</td>
<td><em>sink</em> <em>zinc</em></td>
</tr>
<tr>
<td>High</td>
<td>Coronal</td>
<td><em>tip</em> <em>dip</em></td>
</tr>
<tr>
<td>High</td>
<td>Coronal</td>
<td><em>tomb</em> <em>doom</em></td>
</tr>
<tr>
<td>High</td>
<td>Non-Coronal</td>
<td><em>pit</em> <em>bit</em></td>
</tr>
<tr>
<td>High</td>
<td>Non-Coronal</td>
<td><em>feel</em> <em>veal</em></td>
</tr>
<tr>
<td>High</td>
<td>Non-Coronal</td>
<td><em>could</em> <em>good</em></td>
</tr>
<tr>
<td>Mid</td>
<td>Coronal</td>
<td><em>tore</em> <em>door</em></td>
</tr>
<tr>
<td>Mid</td>
<td>Coronal</td>
<td><em>tent</em> <em>dent</em></td>
</tr>
<tr>
<td>Mid</td>
<td>Coronal</td>
<td><em>toe</em> <em>dough</em></td>
</tr>
<tr>
<td>Mid</td>
<td>Non-Coronal</td>
<td><em>cover</em> <em>govern</em></td>
</tr>
<tr>
<td>Mid</td>
<td>Non-Coronal</td>
<td><em>pair</em> <em>bear</em></td>
</tr>
<tr>
<td>Mid</td>
<td>Non-Coronal</td>
<td><em>face</em> <em>vase</em></td>
</tr>
<tr>
<td>Low</td>
<td>Coronal</td>
<td><em>tangle</em> <em>dangle</em></td>
</tr>
<tr>
<td>Low</td>
<td>Coronal</td>
<td><em>taught</em> <em>dot</em></td>
</tr>
<tr>
<td>Low</td>
<td>Coronal</td>
<td><em>tiny</em> <em>diner</em></td>
</tr>
<tr>
<td>Low</td>
<td>Non-Coronal</td>
<td><em>cotton</em> <em>gotten</em></td>
</tr>
<tr>
<td>Low</td>
<td>Non-Coronal</td>
<td><em>patch</em> <em>batch</em></td>
</tr>
<tr>
<td>Low</td>
<td>Non-Coronal</td>
<td><em>fan</em> <em>van</em></td>
</tr>
</tbody>
</table>
English, practice items

badge
buddy
butter
clubhouse
eagle
powder
ribbon
ripple
sock
supper
wrap
Appendix E  Usable and cutoff-limited duration measurements

The table presents mean numbers of duration measurements deemed usable, per participant, as a function of segmentation region and stimulus type, for two groups of participants defined by language history. Percentages of “outliers” (values subject to speaker-based limiting values) appear in parentheses.

For further detail and discussion, refer to Table 2.2, Table 2.3, and accompanying text in Chapter 2.

<table>
<thead>
<tr>
<th>Region 3 (C1)</th>
<th>Region 4 (V1)</th>
<th>Region 5 (Closure)</th>
<th>Region 6 (C2)</th>
<th>Region 7 (V2…#)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Foreign-born participants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian (N=72)</td>
<td>68.1 (5.5%)</td>
<td>68.1 (2.1%)</td>
<td>68.1 (4.8%)</td>
<td>68.1 (5.5%)</td>
</tr>
<tr>
<td>English, monosyllabic (N=72)</td>
<td>55.0 (6.8%)</td>
<td>55.0 (3.4%)</td>
<td>55.0 (3.4%)</td>
<td>52.9 (3.5%)</td>
</tr>
<tr>
<td>English, multisyllabic (N=48)</td>
<td>38.7 (4.1%)</td>
<td>38.7 (5.5%)</td>
<td>38.7 (3.7%)</td>
<td>38.7 (4.8%)</td>
</tr>
<tr>
<td><strong>U.S.-born participants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian (N=72)</td>
<td>60.3 (3.9%)</td>
<td>60.3 (3.9%)</td>
<td>60.3 (2.2%)</td>
<td>60.3 (6.1%)</td>
</tr>
<tr>
<td>English, monosyllabic (N=72)</td>
<td>68.7 (8.0%)</td>
<td>68.7 (3.6%)</td>
<td>68.7 (3.2%)</td>
<td>62.8 (4.2%)</td>
</tr>
<tr>
<td>English, multisyllabic (N=48)</td>
<td>46.8 (5.3%)</td>
<td>46.8 (2.1%)</td>
<td>46.8 (1.8%)</td>
<td>46.8 (5.3%)</td>
</tr>
</tbody>
</table>

*Note.* There was no Region 7 for English monosyllabic stimuli, where Region 6 completed the target word.
Appendix F  Intrinsic duration and height

The figures below present vowel duration means in milliseconds as a function of language of utterance, target vowel height and [voice] value in the post-vocalic environment.

**Foreign-born speakers**

![Graph showing vowel duration means in milliseconds as a function of language of utterance, target vowel height and [voice] value in the post-vocalic environment for Italian and English utterances.]

**U.S.-born speakers**

![Graph showing vowel duration means in milliseconds as a function of language of utterance, target vowel height and [voice] value in the post-vocalic environment for Italian and English utterances.]

*Note.* The mid vowels for the English words appear to stray from the general pattern. This seeming inconsistency arises from the nature of the stimuli, where the English mid-height vowels are predominantly short vowels.
Appendix G  Italian stimulus sets

The tables present vowel duration means (and standard deviations) in milliseconds as a function of [voice] value of the trigger consonant for the full, 18-pair set of Italian stimuli as found with the core cases and the 6-pair subset as found with the multisyllabic cases. The top table lists values for the foreign-born speaker group and the bottom table lists values for the U.S.-born speaker group.

For further detail, refer to Section 3.2 in Chapter 3.

<table>
<thead>
<tr>
<th></th>
<th>Voiced</th>
<th>Voiceless</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-pair set</td>
<td>186 (38.3)</td>
<td>169 (32.4)</td>
</tr>
<tr>
<td>6-pair subset</td>
<td>191 (52.8)</td>
<td>174 (42.4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Voiced</th>
<th>Voiceless</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-pair set</td>
<td>179 (32.5)</td>
<td>130 (23.2)</td>
</tr>
<tr>
<td>6-pair subset</td>
<td>183 (42.5)</td>
<td>137 (33.3)</td>
</tr>
</tbody>
</table>
Appendix H  Individual flapping data

The tables below list percentages indicating frequency with which each speaker flapped and corresponding counts based on segment surface realization for all multisyllabic English stimuli in which the trigger consonant was underlyingly /t/. For each speaker, there were 12 tokens possible. The top table lists values for the foreign-born speaker group and the bottom table lists values for the U.S.-born speaker group.

For further detail, refer to Section 3.4 in Chapter 3.

<table>
<thead>
<tr>
<th>Foreign-born speakers</th>
<th></th>
<th></th>
<th>Percentage flapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>[t]</td>
<td>[ɾ]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaker 104</td>
<td>10</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Speaker 105</td>
<td>9</td>
<td>1</td>
<td>10%</td>
</tr>
<tr>
<td>Speaker 107</td>
<td>3</td>
<td>4</td>
<td>57%</td>
</tr>
<tr>
<td>Speaker 111</td>
<td>8</td>
<td>2</td>
<td>20%</td>
</tr>
<tr>
<td>Speaker 112</td>
<td>5</td>
<td>6</td>
<td>55%</td>
</tr>
<tr>
<td>Speaker 113</td>
<td>2</td>
<td>8</td>
<td>80%</td>
</tr>
<tr>
<td>Speaker 114</td>
<td>1</td>
<td>11</td>
<td>92%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>U.S.-born speakers</th>
<th></th>
<th></th>
<th>Percentage flapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>[t]</td>
<td>[ɾ]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaker 152</td>
<td>10</td>
<td>2</td>
<td>17%</td>
</tr>
<tr>
<td>Speaker 153</td>
<td>0</td>
<td>12</td>
<td>100%</td>
</tr>
<tr>
<td>Speaker 154</td>
<td>2</td>
<td>10</td>
<td>83%</td>
</tr>
<tr>
<td>Speaker 155</td>
<td>10</td>
<td>2</td>
<td>17%</td>
</tr>
<tr>
<td>Speaker 156</td>
<td>0</td>
<td>11</td>
<td>100%</td>
</tr>
<tr>
<td>Speaker 157</td>
<td>0</td>
<td>11</td>
<td>100%</td>
</tr>
</tbody>
</table>
References


