EFFECTS OF FIRST LANGUAGE VOICING RULES ON THE PERCEPTION AND PRODUCTION OF ENGLISH OBSTRUENT SEQEUENCES BY ADULT HUNGARIAN AND POLISH LEARNERS OF ENGLISH

By

MARISA A. MONTELEONE

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

2009

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

Date	Winifred Strange, Ph.D. Chair of Examining Committee
Date	Robert W. Fiengo, Ph.D. Executive Officer

Supervisory Committee:

Winifred Strange, Ph.D.

Gita Martohardjono, Ph.D.

Robert M. Vago, Ph.D.

THE CITY UNIVERSITY OF NEW YORK

Abstract

EFFECTS OF FIRST LANGUAGE VOICING RULES ON THE PERCEPTION AND PRODUCTION OF ENGLISH OBSTRUENT SEQEUENCES BY ADULT HUNGARIAN AND POLISH LEARNERS OF ENGLISH

by

Marisa A. Monteleone

Adviser: Professor Winifred Strange

The present study explored difficulties in the acquisition of a second language (L2) phonology, looking specifically at the role of native (L1) voicing rules on L2 perception and production. Hungarian and Polish late learners of English performed production and perception tasks with English voicing contrasts in contexts where Hungarian and Polish voicing rules might interfere. American English speakers also participated, for comparison. Each participant produced sentences containing fictional names with obstruent sequences crossing a word boundary (e.g. *I met Gus Barker today*). The non-native participants did show evidence of transfer of their native regressive voicing assimilation rules to their productions of English word-final obstruents, although regressive *devoicing* was observed more often than regressive voicing. Each participant also performed identification tasks with similar sentences (e.g. *I met Jess Geller today*): a four-choice task containing the entire obstruent sequence, and a two-choice task containing sentences in which either the first or last name had been replaced with silence (*e.g. I met Jess [silence] today* or *I met [silence] Geller today*). For word-final obstruents, the non-native listeners were significantly less accurate in the two-choice task than the American English controls, but not significantly different from each other. In the four-choice task, both groups became even less accurate, with the Polish listeners showing a more severe effect than the Hungarian listeners. Overall, there was a slight significant correlation of word-final perception and production scores. For word-initial stops, perception was highly accurate for all groups, with the exception of voiced stops in a voiceless-voiced context. Perception of word-initial /s/ was unexpectedly poor for all three language groups. The pattern of results observed in this study suggests that both L1 phonetic and phonological interference affects perception and production in an L2. Implications for current theoretical models of second language phonology are discussed.

Acknowledgments

This dissertation would not have been possible without the encouragement and support of many individuals. First of all, I would like to thank my dissertation adviser, Dr. Winifred Strange, whose faith and confidence in my abilities gave me the courage to persevere with this project. She has been an ideal mentor, and I am grateful for her dedicated, skillful guidance and infectious enthusiasm for research.

I am also grateful for the support of the members of my supervisory committee. I thank Dr. Gita Martohardjono for her guidance during my initial explorations of second language phonology, and for her thoughtful comments throughout each stage of this research project. I thank Robert M. Vago for his helpful comments and engaging discussions, as well as his wealth of knowledge about Hungarian.

I would also like to thank Dr. Charles E. Cairns and Dr. Virginia Valian, who provided mentorship and invaluable training in my early days at the Graduate Center.

I feel very fortunate to have been a part of the Speech Acoustics and Perception Laboratory, a vibrant group of individuals that fosters creative collaboration and a sense of community. I am indebted to its faculty and student members for their ideas, friendship, and moral support.

I would like to express deep appreciation to my friends and family, whose love and unconditional support have been a constant source of strength. I especially thank my mother and father, John and Pat Monteleone, as well as my grandmother and late grandfather, Juanita and Benjamin L. Abat, for always encouraging my pursuit of higher education. And finally, I thank my husband Steve, who knows more than anyone what a

challenge this dissertation has been, and has never faltered in his commitment to helping me see it through to the end.

Table of Contents

Abstract iv

Acknowledgments vi

Table of Contents viii

List of Tables xi

List of Figures xiii

Chapter 1. Introduction 1

  1.1. Theoretical approaches 3

  1.2. Phonetics and phonology of voicing in English, Hungarian, and Polish 6

    1.2.1. Word-initial voicing 7

    1.2.2. Word-final devoicing 8

    1.2.3. Regressive voicing assimilation 9

    1.2.4. Summary of voicing in English, Hungarian, and Polish 12

  1.3. Second language perception of voicing 12

    1.3.1 Word-initial second language voicing perception 12

    1.3.2 Word-final second language voicing perception 14

    1.3.3. Effect of assimilatory processes on lexical access 17

  1.4. The present study: questions and predictions 21

Chapter 2. Method 28

  2.1. Participants 28

  2.2. Perception Task 29

2.2.1. Stimulus materials 29

2.2.2. Procedures 31

2.2.3. Acoustic analysis of stimuli 32

2.3. Production task 37

2.3.1. Stimulus materials and procedure 37

Chapter 3. Results 39

3.1. Production task 39

3.1.1. Data analysis 39

3.1.2. Method of analysis – non-parametric statistics 41

3.1.3. Word-final fricative voicing and duration 41

3.1.4. Word-final voicing – stops (stop-stop context) 44

3.1.5. Word-final voicing – stops (stop-fricative context) 47

3.1.6. Word-initial fricative voicing and duration 49

3.1.7 Other cues 51

3.2. Perception task 52

3.2.1. Data analysis 52

3.2.2. Word-final perception in the two- and four-choice tasks: AE versus HU & PS groups 52

3.2.3. Word-final perception in the two- and four-choice tasks: HU versus PS groups 55

3.2.4. Effect of voicing context on word-final perception 56

3.2.5 Effect of familiarity of first name 58

    3.2.6. Word-initial stop perception 59
    3.2.7. Word-initial fricative perception 60
  3.3. Perception-Production Relation 61

Chapter 4. Discussion 63
  4.1. Summary of results 63
    4.1.1. Word-final production 63
    4.1.2. Word-initial production 65
    4.1.3. Word-final perception: two-choice versus four-choice task 66
    4.1.4. Word-final perception across voicing contexts 70
    4.1.5. Word-initial stop perception 73
    4.1.6. Word-initial fricative perception 74
  4.2. Perception-Production relation 76
  4.3. Implications for theoretical models 76
  4.4. Conclusions and directions for future research 79

5. Tables 82

6. Figures 100

7. Appendices 107

8. References 132

List of Tables

Table 1.1: Summary of voicing in English, Hungarian, and Polish 82

Table 2.1: Perception stimuli, broken down by sequence type, word-initial sequence, and word-final sequence 83

Table 2.2: Summary of acoustic measurement data for perception stimuli – word final consonants 84

Table 2.3: Summary of acoustic measurement data for perception stimuli – word-initial stops 85

Table 2.4: Summary of acoustic measurement data for perception stimuli – word-initial fricatives 86

Table 3.1: Word-final percent voicing across languages – production data 87

Table 3.2: Median word-final fricative duration across language – production data 88

Table 3.3: Median word-initial percent voicing across languages – production data 89

Table 3.4: Median word-initial fricative duration across languages – production data 90

Table 3.5: Median vowel durations across languages 91

Table 3.6: Median VOT durations across languages 92

Table 3.7: Median word-final percent accuracy before an inserted pause – two-choice task 93

Table 3.8: Median word-final percent accuracy before an obstruent – 94

four-choice task

Table 3.9: Overall median word-final percent accuracy across voicing contexts – four-choice task 95

Table 3.10: Word-final percent accuracy broken down by name 96

Table 3.11: Median word-initial stop percent accuracy after an inserted pause – two-choice task 97

Table 3.12: Median word-initial stop percent accuracy after an inserted pause – four-choice task 98

Table 3.13: Median word-initial fricative percent accuracy 99

List of Figures

Figure 2.1: Waveform and spectrogram – fricative-stop sequence 100

Figure 2.2: Waveform and spectrogram – stop-stop sequence 101

Figure 2.3: Waveform and spectrogram – stop-fricative sequence 102

Figure 3.2: Percent word-final correct overall before an inserted pause – two-choice task 103

Figure 3.3: Percent word-final correct overall before an obstruent – four-choice task 104

Figure 3.4: Scatterplot illustrating the relationship between non-native perception and production scores. 105

Chapter 1. Introduction

Speech perception research over the last several decades has shown that adult learners of a second language (L2) often have difficulty identifying and discriminating non-native speech sounds. Early research focused on the perception of single segment contrasts in specific contexts, but more recently, attention is being given to the perception of sequences (e.g. Hallé, Segui, Frauenfelder, & Meunier, 1998; Dupoux, Kakehi, Hirose, Pallier, and Mehler, 1999) and perception across different contexts (e.g. Strange, Akahane-Yamada, Kubo, Trent, and Nishi, 2001; Levy, 2009). Attention to L2 performance in domains larger than the segment is beneficial, not only because it is more representative of real world conditions, but also because it allows us to observe how language-specific interactions between adjacent speech sounds affect L2 perception. Language-specific assimilatory processes, such as voicing and place assimilation, operate in domains larger than a single segment. In the lexical processing literature, these assimilatory processes have been shown to affect lexical access in one's own language (e.g. Gaskell and Marslen-Wilson, 1998; Gow, 2001, 2003; Otake, Yoneyama, Cutler, and van der Lugt, 1996) and, more recently, in a second language (Darcy, Peperkamp, and Dupoux, 2005). These studies have shown that listeners compensate for phonetic alternations due to assimilatory processes when making decisions about lexical items. It seems likely, then, that assimilatory processes affect perception of individual sounds in both the first language (L1) and second language. The current study attempts to explore the effect that an L1 voicing assimilation rule has on the L2 perception of contrasts in a language that does not have the same rule.

Both Hungarian and Polish contain the phonological rule of regressive voicing assimilation (RVA), which neutralizes word-final voicing contrasts before obstruents. In regressive voicing assimilation, the voicing of the rightmost member of an obstruent sequence spreads to the preceding segment. This L1 rule has been shown to transfer to the L2 productions of word-final consonants in English by both native Hungarian and native Polish speakers. In an error analysis of the speech of two native Hungarian late learners of English, Altenberg and Vago (1983) found instances of both regressive devoicing (e.g. *Lookin[k] for*; *beyon[t] his*) and regressive voicing (e.g. *Abou[d] the*; *i[v] the*). Rubach (1984) observed the same pattern of transfer for Polish learners of English, in phrases such as *the[s]e people* and *Thi[z] boy*. Given that the regressive voicing rule creates errors in English L2 productions, the question remains as to whether it will affect L2 perception, as well.

In order to explore the role that this L1 rule plays in L2 perception, the present experiment assessed the production and perception of English obstruent sequences by native Hungarian and Polish late learners of English. Hungarian and Polish serve as interesting test cases for this type of research, since they differ with respect to voicing in pre-pausal contexts. Hungarian maintains a word-final voicing contrast before a pause, but Polish does not, due to a final obstruent devoicing rule. The difference in voicing rules across the two languages should help to tease apart various sources of L1 interference in L2 perception and production.

The details of the experimental design are discussed in Chapter 2. The remainder of this chapter will present a brief discussion of the relevant background literature, followed by a description of specific research questions and predictions.

1.1. Theoretical approaches

This section will summarize a few of the most prominent approaches to second language phonological acquisition. These approaches attempt to predict and explain the difficulties that adult L2 learners face when acquiring a second language.

An early approach to second language phonology, the Contrastive Analysis Hypothesis (CAH; Lado, 1957), asserted that areas of difficulty in second language learning could be predicted through a comparison of the similarities and differences between the native and target languages. Structures that differ between the native and target language should be difficult to acquire, and those that are similar should be easier. While this approach does indeed predict some areas of difficulty, Eckman (1977, 2008) showed that the CAH does not predict the *direction* of difficulty. For example, the CAH would predict that the presence of RVA in Polish and the lack of RVA in English might cause problems for second language learners. The CAH does not, however, predict if this difference will cause a problem for Polish speakers learning English, or for English speakers learning Polish. In order to remedy the CAH, Eckman proposed the Markedness Differential Hypothesis (MDH). The MDH starts from the basic premise that differing structures between the native and target languages should be difficult to acquire, and includes the concept of *markedness* to account for the direction of difficulty. The MDH proposes that when two languages differ with respect to a given structure, the language containing the marked structure will cause difficulty for speakers of the language with the unmarked structure. Typically, marked structures are those structures that are more rare cross-linguistically. According to the MDH, given two languages with different structures, the marked structure is the one that occurs in one language, but not the other. Returning to the English/Polish example, in English it is possible to have voicing sequences with a voicing mismatch, but in Polish it is not. The English structure, therefore, is more marked, and therefore predicted to be harder to learn. The example of mispronunciations by Polish speakers mentioned above (*thiz boy*) appears to provide support for this hypothesis.

Both the CAH and MDH are situated in a generative linguistic theoretical approach to language acquisition. A prominent approach to L2 perception and production in the speech science literature is the Speech Learning Model (SLM; Flege 1995). The SLM attempts to account for the age-related difficulties that individuals face when learning to produce the sounds of a second language. Flege claims that while our ability to form phonetic categories remains intact throughout our lifetime, our ability to perceive differences between certain L1 and L2 categories decreases with age of L2 acquisition. The older one is when beginning to learn a second language, the more difficult it is to detect differences between native and non-native categories. The SLM predicts that L2 categories that differ perceptually from existing L1 categories will be acquired more easily than those that are similar. Since, according to this model, the ability to create phonetic categories is retained throughout the lifespan, an L2 learner should be able to create categories for ‘new’ speech sounds. For example, for native Spanish speakers, the English vowel /æ/ may be initially difficult to discriminate from /a/. However, native Spanish speakers who have more experience with English eventually improve on the /a/ - /æ/ contrast (Flege, Munro, and Fox, 1995). Flege suggests that the experienced speakers were able to create a category for /æ/ due to the fact that Spanish lacks /æ/ -- it is a ‘new’ speech sound. Difficulty in L2 category formation arises, however, when learners are unable to discern the difference between an L2 phone and an existing category. Non-native phones that are perceptually similar to existing L1 categories may be linked to the L1 category via a process of equivalence classification, blocking the formation of a new category. Spanish /i/, for example, is acoustically close to English /i/, so it is predicted that native Spanish-speakers will be more likely to establish a new category for English /æ/ than English /i/. Note here the difference between the SLM and the CAH – while the CAH asserts that differences between languages cause difficulty, the SLM claims that significant phonetic differences may help the learner in the L2. That is, if a new category is different enough from existing categories, it may be easier to learn.

Another prominent approach is the Perceptual Assimilation Model (PAM; Best, 1995). Best's model is designed to predict performance in the discrimination of non-native contrasts by naïve listeners. Best recognizes that due to the considerable overlap in the phoneme inventories across languages, non-native contrasts will be perceived according to the way in which they are perceptually assimilated to native categories. Assimilation to native categories depends on the degree of gestural similarity between the native and non-native segments. Successful discrimination of a non-native contrast depends on a) the goodness of fit of each of the L2 phones to L1 categories and b) the respective categories of the members of the contrast.

If the members of a non-native contrast correspond to different categories in the native language, discrimination will be very good. This is referred to as a *two-category assimilation* pattern in PAM. For example, while AE /ɹ/ is phonetically distinct from Spanish /r/, Spanish discrimination of AE /ɹ/ and /l/ is good, since the two members of the contrast assimilate to separate categories in Spanish. Difficulties arise when the two members of the contrast assimilate to the same category, in a *single category assimilation* pattern. In Japanese, both members of the AE /ɹ / - /l/ contrast are assimilated to one phoneme, /R/, and are considered poor exemplars of this category (Guion, Flege, Akahane-Yamada, & Pruitt, 2000). Perception, then, is accurately predicted to be poor. However, if two members of a non-native contrast assimilate to the same category, but one is a better exemplar than the other (in a *category goodness* difference pattern), then assimilation is predicted to be better.

The predictions of PAM are meant to account for perception by naïve listeners of an unfamiliar target language. That is, these naïve listeners are not actively learning the target language. Best and Tyler (2007) set forth a version of PAM which accounts for perception by second language learners – PAM-L2. Whereas PAM predicts the starting point for L2 learners of a language, PAM-L2 makes predictions about the L2 learning process over time. In PAM-L2, contrasts that are predicted by PAM to be successfully discriminated by naïve learners are predicted to continue to be successfully discriminated as an individual actively learns the L2 over time. Discrimination of more difficult contrasts, such as phones that are assimilated to the same native category, may continue to be difficult. This is due to the fact that, like the SLM, PAM-L2 predicts that there will not be much perceptual learning for exemplars which are considered good exemplars for a native category. For example, if the target language /i/ is a good exemplar of the native language /i/, it is not very likely, according to both the SLM and PAM-L2, there will not be much perceptual learning for the target language /i/.

1.2. Phonetics and phonology of voicing in English, Hungarian, and Polish

1.2.1. Word-initial voicing

The concept of *voice onset time* (VOT), first discussed by Lisker and Abramson (1964), is used to describe the differences in the phonetic realization of voicing in stops across languages. VOT is defined as the timing of the onset of voicing (as evidenced by periodic glottal pulsing) relative to the release of stop closure. If voicing begins prior to the release of the stop closure, this is referred to as negative VOT. If voicing begins shortly after the release of the stop closure, as in the unaspirated [p] in the English word *spin,* this is referred to as short-lag VOT. Finally, if the onset of voicing lags due to the presence of aspiration, as it the English word *pat*, this is referred to as long-lag VOT. In Polish and Hungarian, word-initial voiced stops are characterized by negative VOT, while voiceless stops are characterized by short-lag VOT. That is, voiceless word-initial stops are realized with little aspiration. Acoustic evidence supports this characterization (Lisker & Abramson, 1964, Gosy 2001 for Hungarian; Keating, Mikos, & Ganong, 1981 for Polish) noting that the voiceless velar stops have higher VOT values than those of alveolar and bilabial stops. English word-initial stops follow a different pattern than those of Hungarian and Polish. English word-initial voiceless stops are characterized by long-lag VOT, as they are realized with aspiration. Voiced stops may be realized with negative VOT in English. However, they are more commonly realized with short-lag VOT as voiceless unaspirated stops.

English, Hungarian, and Polish all have a word-initial /s/ - /z/ contrast. One potential confounding issue for Hungarian speakers learning English is that the spelling for English /s/ is different to that in Hungarian. English /s/ is written as **sz** in Hungarian, while English /ʃ/ is written as **s** in Hungarian.

1.2.2. Word-final devoicing

In word-final position, Polish has final devoicing of obstruents. That is, word-final voiced stops and fricatives become voiceless before a voiceless segment or a pause. Examples from Rubach (1996, p. 70) are as follows:

Polish Final Devoicing

sa[d] + y ‘orchards’ – sa[t] (nom.sg.)

ko[z] + a ‘goat – kó[s] (gen.pl.)

Whether or not the devoicing rule fully neutralizes the voicing contrast, however, is unclear. Slowiaczek and Dinnsen (1985) claim that the word-final devoicing rule is not fully neutralizing. In an acoustic study of Polish, they found that before neutralized word-final stops and fricatives, vowel durations were longer for underlying voiced than voiceless consonants. A later study by Jassem and Richter (1989), however, noted that the Slowiaczek and Dinnsen study failed to control for certain factors (for example, their participants had been living in an English-speaking country for an unspecified time, the target words were embedded in contexts that arguably may not trigger final devoicing and the method of elicitation – reading from orthographic transcriptions – favored hypercorrect speech). Jassem and Richter replicated the study with the proper controls, and found that there were no significant differences between underlying voiced and voiceless obstruents in final position.

Hungarian, in contrast to Polish, does not have a rule of final obstruent devoicing. Obstruents in word-final position may be voiced or voiceless, as illustrated by the minimal pairs:

Hungarian Final Voicing Contrast

lá[p] 'marsh' – lá[b] 'leg'

fo[k] 'rung' – fo[g] 'tooth'

ké[s] 'ready – ké[z] 'hand'

English also maintains a contrast in word-final position, as illustrated by the minimal pairs *cap/cab* and *rip/rib*. English final obstruents do show some level of word-final devoicing. This devoicing, however, is described as partial (Ladefoged, 1993). Word-final voicing contrasts have been shown to have a number of voicing cues in English (Rapahel, 1981). In an early study, Raphael (1972) claimed that vowel duration was a primary cue for the word-final voicing contrast in English, with longer vowel durations cueing voiced consonants, and shorter durations cueing voiceless consonants. This claim was based on a perception study using synthetic minimal pairs produced in isolation. In a later perception study with natural (edited) stimuli in sentential context, Raphael (1981) retreated somewhat from his earlier position, claiming that there were a large number of cues to the voicing distinction, each of which had some weight. Closure voicing in particular was found to be a consistently effective cue for the voicing distinction, but all cues tested (including vowel duration, release burst, closure duration, and transitions) had some value. An interesting finding from this study relevant to the current study is that the stimulus *peg* was perceived correctly 90% of the time before the voiceless fricative /ʃ/, indicating that American English listeners can perceive word-final voicing even when the following obstruent is voiceless (i.e. when there is a voicing mismatch in a sequence of obstruents).

1.2.3. Regressive voicing assimilation

Hungarian and Polish both contain a rule of regressive voicing assimilation word-internally and across word boundaries (Vago, 1980, Rubach, 1996, Jansen, 2004, Petrova et al., 2006). The following examples adapted from Vago (1980, p. 34) illustrate Hungarian regressive voicing assimilation across morpheme boundaries:

Hungarian Regressive Voicing Assimilation

Stem	Suffix	
kút 'well'	-ban 'in'	kútban[db]
zsák 'sack'	-ban 'in'	zsákban [gb]
rab 'prisoner'	-tól 'from'	rab tól [pt]
meleg 'warm'	-től 'from'	melegtől [kt]

The following examples reproduced from Rubach (1996, p. 71) illustrate Polish regressive voicing assimilation across word boundaries:

Polish Regressive Voicing Assimilation

nos+ i + c [ś] 'carry' – noś + że [ź ž] 'do carry' (imper.)[1]

szlak + u 'route' [k] (gen.sg.) – szlak bojowy [g b] 'war route'

I have not found any acoustic studies of Polish regressive voicing assimilation, but acoustic studies of Hungarian and English are discussed below.

In an acoustic study, Jansen (2004) examined Hungarian productions of /k g ʃ ʒ/ followed by /t d s z l r/. Results showed that voicing (as indicated by duration of periodicity in the waveform and voice bar in the spectrogram) was fully neutralized before voiceless obstruents. Before voiced obstruents, however, while voicing of /k/ and

---

[1] Rubach diverges from IPA symbols in certain instances: [ś ź] – prepalatal fricatives; [ž] – postalveolar fricative

/ʃ/ increased, it did not reach the levels found in their underlyingly voiced counterparts. It appears from these data, then, that regressive *devoicing* seems to be more complete than regressive voicing in this language. When the first consonant of an obstruent-obstruent sequence was a stop, vowel duration differences were reduced or erased, indicating that assimilation is fully neutralizing in these contexts. When the first consonant of an obstruent-obstruent sequence was a fricative, however, a vowel length distinction was maintained. It seems, then, that although acoustic evidence supports the existence of regressive voicing assimilation in Hungarian, the extent to which this process is fully neutralizing is not completely clear.

In the same acoustic study, Jansen (2004; 2007) examined the same consonant sequences in British English as in the Hungarian experiment. Results showed that /t/, /s/, and /z/ caused some phonetic deviation in the voicing of preceding velar stops consistent with regressive voicing assimilation. Vowel durations, however, were not affected before assimilation contexts, suggesting that the assimilation was non-neutralizing. Interestingly, the phoneme /d/ did not trigger regressive assimilation in the same way that /z/ did. While there was some evidence that both lowered F1 values of the preceding vowel, only /z/ triggered longer voicing durations of preceding stops. Jansen suggests that this is due to the fact that /d/, while phonologically 'voiced' is phonetically voiceless with short lag VOT.

Stevens, Blumstein, Glicksman, Burton, & Kurowski (1992) also showed acoustic evidence of regressive voicing assimilation in English. As part of a larger study of fricative voicing, Stevens et al. examined voicing in fricative-fricative sequences with mixed voicing that crossed a word boundary (e.g. [z#f], [v#s]). Voicing, as measured by glottal vibration, was maintained for voiced fricatives in mixed clusters only 24% of the time. Examination of the direction of the assimilation showed regressive devoicing to be greater than progressive devoicing. This finding was confirmed in a perception study with the same data. Listeners, however, were more accurate at identifying underlying voiced fricatives than acoustic measures predicted, leading the researchers to conclude that glottal vibration alone was not sufficient enough as a cue to voicing in mixed fricative sequences.

1.2.4. Summary of voicing in English, Hungarian, and Polish

In summary, English, Hungarian, and Polish all maintain a voicing contrast in word-initial position, but English realizes voicing in word-initial stops differently than Hungarian and Polish. While Hungarian and Polish word-initial stops are distinguished by negative versus short-lag VOT, English stops tend to be distinguished by short-lag versus long-lag VOT. In word-final position, English maintains a voicing contrast, but does show some evidence of regressive voicing assimilation. Hungarian maintains a voicing contrast before a pause, but not before an obstruent, and Polish does not have a word-final voicing contrast in any context. See Table 1.1. for a summary of this information.

1.3. Second language perception of voicing

1.3.1 Word-initial second language voicing perception

Early studies of perception of word-initial voicing looked at the identification and discrimination of synthetic VOT continua (e.g. Abramson & Lisker, 1973, Willams, 1977, Keating, Mikos, & Ganong, 1981, Flege and Eefting, 1986). In these identification tasks, listeners are asked to label a series of stimuli ranging in VOT values (e.g. from -

150 to +150 milliseconds (ms) in the Abramson and Lisker study). These studies show that labeling functions tend to correspond with a speaker's native language category boundaries. For example, Flege and Eefting (1986) showed that native English listeners had a boundary between voiced and voiceless /t/ - /d/ at a higher average VOT value than native Spanish subjects (36.2 ms. vs. 19.9 ms., respectively). This corresponds with production data from the same study showing that English /t/ has long-lag VOT, while Spanish /t/ has short-lag VOT.

The existence of different category boundaries across languages has implications for word-initial voicing perception in a second language. A speaker of a language with a category boundary between /b/ and /p/ at +5 milliseconds may interpret short-lag [b̥] in English as /p/. English [p^h], with a long-lag positive VOT, however, should not be problematic for a speaker with such a boundary, since it would still fall into the voiceless category.

A study by Keating, Mikoś, & Ganong (1981), however, showed that the English boundary may not necessarily cause difficulty for non-native speakers. The study observed the labeling functions of a synthetic /d/ - /t/ continuum across different ranges of VOT by speakers of Polish and English. They found that labeling functions of /d/ and /t/ were consistent for English speakers, even when the range of the VOT continuum was manipulated (-100/+20, -100/+50, and -20/+80). That is, the median VOT boundaries were consistently between about +28 and +46, regardless of whether the listener was presented with stimuli over a -100 to +20 millisecond range or a -20 to +80 millisecond range. Labeling functions of native Polish speakers, however, shifted across the different VOT ranges. That is, their median VOT boundary was +5 and +6 for the ranges including several negative VOT values (-100/+20 and -100/+50, respectively), but +20 for the -20/+80 range. This finding was replicated with other ranges in other experiments within the same study. This suggests that the English category boundary might somehow be salient even for those with a different boundary in their native language.

The Keating, Mikoś, & Ganong study, however, was performed with synthetic stimuli, so it is difficult to know whether or not her findings would extend to the perception of natural speech. A study of native Italian bilinguals explored the perception and production of the English voicing contrast using natural stimuli (MacKay, Flege, Piske, and Schirru, 2001). Italian is like Polish and Hungarian in that it uses the negative vs. short-lag VOT categories for its voicing contrast. The majority of the study focused on the production of the English stops /b, d, g, p, t, k/ by Italians with varying levels of experience with English, but it is the results of the perception study that are of interest here. Italian late bilinguals (who were first exposed to English in adolescence or adulthood) misidentified word-initial /b, d, g/ tokens as /p, t, k/ more than Italian early bilinguals and native English speakers. Errors in the identification of /p, t, k/ by Italians did not significantly differ from native English speakers.

1.3.2 Word-final second language voicing perception

Studies on word-final perception of voicing show that speakers of languages that lack a voicing contrast in final position learn the English word-final contrasts with varying degrees of success. In experimental studies of English word-final voicing perception by non-native listeners, accuracy rates have been shown to vary based on the strength of the acoustic cues present in the stimuli. In a study of Mandarin late learners of English, Flege (1989) showed that the presence of release bursts positively affected word-final /t/ - /d/ perception in a forced choice task. Although Mandarin does not have consonants in word-final position, participants were able to identify the English word-final /t/ - /d/ contrast with high accuracy rates when presented in isolated syllables with the release burst and closure voicing cues present. When the release burst and closure voicing cues were removed, however, accuracy rates dropped significantly. Native English control subjects had no problem with these same stimuli. Since English word-final stops are likely to lack release burst cues in conversational speech, Flege concluded that Mandarin speakers might have difficulty with this contrast in everyday usage of English.

A later study by Flege and Wang (1989) provided clues to why Mandarin speakers might have problems with this particular contrast. They again looked at native Mandarin speakers' perception of /t/ - /d/ with the burst and closure voicing cues removed, but in addition they also looked at the performance of native Shanghainese and native Cantonese late learners of English. Native Shanghainese speakers were predicted to do better than the Mandarin speakers because Shanghainese does have one word-final consonant: a glottal stop. The native Cantonese speakers were predicted to do better still, since Cantonese has word-final voiceless stops and nasals. Flege and Wang predicted that since the Shanghainese and Cantonese speakers had more experience attending to word-final vowel transition cues, they should be better prepared to perceive the English word-final voicing contrast. Their predictions were confirmed: Shanghainese and Cantonese speakers had higher identification rates than the Mandarin speakers on the contrast, with native Cantonese speakers demonstrating the highest identification rates. This suggests that experience attending to the end of words aids in the detection of non-native word-final contrasts.

Experience with word final consonants in one's L1, however, does not necessarily indicate that the English word-final voicing contrast will be easy. Pikser (2003) examined the perception of voicing and manner contrasts among word-final alveolar consonants /t d s z/ in a same-different task with L2 learners whose native language was Spanish. Spanish does have word-final consonants, but does not have a voicing contrast in word-final position. She found that while manner distinctions were discriminated with high accuracy, voicing distinctions were met with more difficulty, especially for stop consonants. Similar to the results from Flege (1989) and Flege and Wang (1989), the strength of the release burst cue had an effect on perception. Of the two speakers that produced the stimuli, one had significantly longer release bursts. Listeners were more accurate in discriminating the voicing contrasts in word-final stop consonants produced by this speaker.

Even when L2 learners of English perceive word-final voicing contrasts with high accuracy levels, experimental evidence shows that they attend to different cues than native English speakers. Broersma (2005) showed that Dutch listeners who were highly proficient in English were able to identify word-final voicing contrasts in isolated nonwords as accurately as native English speakers. The accuracy levels remained high even for final stops which had the release bursts removed. Broersma attributes the success of the Dutch speakers to their experience with the voicing contrast in word-initial and word-medial position, as well as their experience attending to a vowel duration cue. A second experiment in Broersma (2005) showed, however, that Dutch listeners do not use cues to final voicing in the same way as native English speakers. Participants were asked to label edited natural tokens along a continuum (from /ku:v/ to /ku:f/, for example). English native speakers used vowel duration differences to determine the identity of the final labiodental fricative more than the native Dutch speakers.

This finding is consistent with findings from Flege and Hillenbrand (1986) who examined the identification of the American English word-final /s/ - /z/ distinction with a forced choice identification task using a synthetic *peace – peas* continuum. Although native speakers of French, Swedish, and Finnish could all discriminate the contrast at the ends of the continuum, each attended to cues consistent with their native language. Native French listeners used the vowel duration cue less than American English listeners, which was consistent with their production data. The Swedish and Finnish listeners attended to the vowel duration cue *more* than the American English listeners, perhaps because Swedish and Finnish both have phonemic vowel length distinctions, and therefore had experience attending to such a cue. What is consistent in both the Broersma and Flege and Hillenbrand studies is that non-native listeners attend to different acoustic cues than do native English speakers, even when they perceive contrasts correctly.

1.3.3. Effect of assimilatory processes on lexical access

In order to determine how L1 assimilation rules might interfere with L2 perception, it is useful first to ask how assimilation affects L1 perception. Studies from the lexical processing literature have shown that in most cases phonetic variability due to assimilation does not hinder processing by native listeners. Gaskell and Marslen-Wilson (1996) explored whether phonetic variation due to place assimilation would inhibit lexical access. In English, coronal segments often assimilate to the place of articulation of a following consonant (as in *wicked prank* → *wicke[b] prank*). Using a cross-modal repetition priming task, they showed that changed versions of target words, such as *wicke[b]*, did not negatively affect lexical access as long as the change was phonologically viable (e.g. *wicke[b] prank*). When changed words were presented in isolation or in phonologically unviable contexts (e.g. *wicke[b] game*), access of the target word was inhibited. Similarly, Otake, Yoneyama, Cutler, and van der Lugt (1996), showed that a Japanese place assimilation rule did not inhibit processing of the moraic nasal /n/ by Japanese listeners. In a phoneme monitoring task using Japanese words, all allophonic variants of the moraic nasal yielded similar response times and accuracy rates. That is, assimilated [m] in *tonbo*, [ŋ] in *rongo*, and [n] in *konro* were all perceived equally as /n/. Subsequent experiments within the same Otake et al. study went on to show that not only is phonological variation due to place assimilation well-tolerated in Japanese, it also provides information that may facilitate processing. Another phoneme-monitoring task with cross-spliced stimuli showed that nasal-consonant sequences with the same place of articulation (e.g. *de[mp]a*) resulted in faster detection of /n/ than tokens that violated assimilation rules (e.g. *ko[m]to*). Processing of the *following* consonant was also facilitated when the nasal-consonant sequence shared the same place of articulation. Accuracy rates and reaction times for the consonant following the nasal (e.g. the /b/ in *tonbo*) were significantly improved when the place of articulation of the nasal matched (e.g *to[mb]o*) rather than mismatched (e.g. *to[nb]o*).

It appears, then that listeners use cues from the assimilated segment in anticipating the place of articulation of the following segment. This is confirmed in other studies (e.g. Gaskell and Marslen-Wilson, 1998; Gow, 2001, 2003). The finding that following contexts facilitate the processing of the identity of an assimilated segment itself is also confirmed in other studies (e.g. Gaskell and Marslen-Wilson, 1998, Gow 2003). But certain factors make recovery of an assimilated segment more difficult. Gaskell and Marslen-Wilson (2001) showed that if a word that has undergone assimilation is lexically ambiguous, it may not prime the original form without a biasing semantic context. Their study focused on words such as *run*, where the assimilated form is also a real word. In the case of the word *run*, the assimilated form due to labial place assimilation, *ru[m]*, is also a real word. In a cross-modal priming task, Gaskell and Marslen-Wilson used word pairs such as *rum/run*. Forms such as *rum* did not prime their coronal counterparts (*run*), even in a phonologically viable context (e.g. "A quick *rum* picks you up"). Only when primes were inserted in a biasing semantic context which favored the coronal counterpart (e.g. "It's best to start the day with a burst of activity. A quick *rum* picks you up."), did labial-final forms prime a coronal-final target. Gaskell and Marslen-Wilson's findings suggest that it may be difficult to access the underlying form of lexically ambiguous, fully assimilated utterances. Gow (2002), however, in a cross-modal priming study using naturally assimilated tokens, showed evidence that place assimilation in English did not create lexical ambiguity, even when tokens were fully assimilated (according to offline judgments). Assimilated coronals such as the /t/ in *ri[p]e berries* primed coronal targets (e.g. *right*) more than their labial counterparts (e.g. *ripe*). Gow suggests that acoustic features of underlying forms are still preserved in the surface forms of assimilated utterances, and these features are available to listeners. So, while the utterances used in the priming study were rated in offline judgments as being highly assimilated, they were not as fully assimilated as the homophones used in the Gaskell and Marslen-Wilson study. Since the stimuli in the Gow study were naturally produced, his results may more closely reflect what listeners do in a conversation in a natural listening environment.

It seems from the evidence given above, that native speakers are accustomed to dealing with assimilation in their own language, and even use this information to make decisions about adjacent segments. This process is unproblematic as long as targets are not lexically ambiguous. When targets are lexically ambiguous, access of the underlying form may be more challenging. If assimilation is complete, listeners may need to rely on semantic cues to access the underlying form. If assimilation is incomplete, listeners may be able to use acoustic cues to detect the underlying form.

Given that listeners appear to take assimilation rules into account when processing lexical items in their native language, the question remains if they will continue to do so when processing other languages. Evidence from a study by Darcy, Peperkamp, and Dupoux (2005) suggests that listeners do carry over their native assimilation rules to processing of a second language. They examined native and non-native processing of voice and place assimilation rules by native French and native English listeners. French has a voicing assimilation rule, but not a place assimilation rule, while English has a place assimilation rule, but no voicing assimilation rule. The researchers found that French speakers compensated for voicing assimilation, but not for place assimilation, when listening to French sentences. For example, the form *ro[p]* was identified as the target *robe* more often when it was in a context consistent with voicing assimilation (e.g. *ro[p] sale*) than one that was not (e.g. *ro[p] noire*), but the form l*u[m]* was equally rejected as the target *lune* whether it was in a context consistent with place assimilation (e.g. *lu[m] pale*) or one that was not (e.g. *lu[m] russe*). English speakers were found to do the opposite in their native language, compensating more for place assimilation than voicing assimilation. Interestingly, when participants were tested in their second language, they tended to continue to compensate for native rules rather than the non-native rules. Beginning English learners of French continued to compensate more for place assimilation than voicing assimilation when listening to French sentences, while beginning French learners of English compensated more for voicing than place assimilation when listening to English. Advanced L2 learners began to pattern more like native speakers, compensating less for their L1 rule, and more for the L2 rule. For the purposes of this study, an important note is that both English and French speakers compensated for regressive *devoicing* in English (as in *bi[k] [f]ountain*). Darcy et al. suggest that this is due to the partial phonetic devoicing observed in American English in this context.

In summary, the evidence from studies of the perception of assimilation in the lexical processing literature suggests that native assimilation rules do affect the processing of lexical items in the L1, and continue to do so even in the L2. The finding that voicing assimilation rules can influence lexical decisions in the L2 is particularly interesting, since we know from the second language speech perception literature that the perception of voicing in an L2 can be challenging. More research is needed that fills in the gap between the second perception literature on the one hand, and the lexical processing literature on the other. Are native voicing assimilation rules one of the contributing factors to perception difficulties in the L2?

1.4. The present study: questions and predictions

The present study examined the perception and production of English obstruent sequences by late learners of English whose native languages were Hungarian and Polish. Hungarian, Polish, and American English participants identified word-final and word-initial consonants in fictional names containing an obstruent sequence crossing a word boundary (e.g. *I met Jess Geller today*). Since these American English sequences are in a context where regressive voicing assimilation occurs in both Hungarian and Polish, it was predicted that the non-native listeners would have difficulties with this perception task. That is, the L1 rule is predicted to interfere with L2 perception. Word-final and word-initial perception were assessed separately. Word-final obstruents were identified both in the assimilation context (e.g. *Jess Geller*) and before an artificially inserted pause that was substituted for the acoustic segments that constituted the last name (e.g. *I met Jess [silence] today*). Word-initial obstruents were identified both in the assimilation context (e.g. *Jess Geller*) and following an artificially inserted pause that was substituted for the acoustic segments that constituted the first name (e.g. *I met [silence] Geller today*). These additional edited sentences were included to determine whether perceptual errors were due to the acoustic properties of the individual segments or the assimilation context itself. The same exact utterances were used in the pre- and post-obstruent conditions as the pre- and post-pausal conditions. After the perception task, all participants subsequently produced sentences containing similar fictional names with obstruent sequences crossing a word boundary.

The prediction that L1 voicing rules would interfere with the perception of these L2 sequences assumes that L2 perception makes use of context-dependent information which is integrated at a higher, abstract phonological level. This is counter to the view that interference occurs at a language-specific phonetic level where listeners incorrectly map acoustic cues to L1 categories independent of context. For convenience, throughout the remainder of the discussion the former account will be referred to as an L1 phonological interference hypothesis, and the latter as an L1 phonetic interference hypothesis. The two forms of L1 interference are not mutually exclusive, and it is likely that both affect L2 speech perception to some degree. But the patterns of errors resulting from each of these two different types of interference are predicted to be quite different. For example, if non-native perception of American English word-final /s/ is solely due to the way in which the acoustic cues to the American /s/ are mapped onto Hungarian (or Polish) /s/, then perception in the pre-obstruent and pre-pausal conditions would be expected to be the same. However, if listeners are integrating information about the entire sequence, then perception of word-final /s/ might be different if /s/ is followed by a voiced versus voiceless sound, then perception in the pre-obstruent and pre-pausal conditions might be different. Predictions based on the two hypotheses are outlined below. It was expected that neither of these 'pure' sets of predictions would account for all the error patterns, and that both types of interference would have an effect on the results. But setting up predictions in this way allows us to tease apart the relative roles of each type of interference.

As stated above, the present experiment consisted of both perception and production tasks. The research questions and predictions are outlined below.

Question 1: Do native Hungarian and Polish speakers transfer their L1 voicing patterns to their L2 productions of English?

Predictions:

Word-final voicing: If native Hungarian and Polish speakers transfer their L1 regressive voicing assimilation into their L2 productions of English, we should expect the voicing of the second consonant (word-initial) of an obstruent sequence to dictate the voicing of the entire sequence. If the word-initial segment is voiced, the preceding word-final segment should be voiced, and if the word-initial segment is voiceless, the preceding word-final segment should be voiceless, regardless of underlying voicing.

Based on evidence from similar studies, it is likely that regressive *devoicing* will be observed more than regressive voicing. As discussed above, Jansen (2004) found that regressive devoicing was more prevalent than regressive voicing across word boundaries in Hungarian. Similarly, in an L2 production study, Cebrian (2000) showed that Catalan L2 learners of English were more likely to devoice word-final obstruents that preceded voiceless segments than they were to voice obstruents that preceded voiced segments.

Word-initial voicing: If Hungarian and Polish speakers transfer their native voicing categories into their English productions, they should maintain a voicing distinction in word-initial position even when preceded by a word-final obstruent. The VOT values of their word-initial stops, however, should resemble those of native categories. That is, voiced stops should be characterized by negative VOT, and voiceless stops by short-lag, positive VOT.

<u>Question 2</u>: How will Hungarian and Polish listeners perceive English word-final voicing contrasts?

a. Will the Hungarian participants perform more accurately than the Polish participants, due to the fact that Hungarian has a word-final voicing contrast before a pause?

b. Will Hungarian and Polish listeners identify the contrasts less accurately when followed by an obstruent versus an inserted pause?
c. Will some voicing contexts be more difficult than others?

Predictions:

Since Hungarian has a word-final voicing contrast before a pause, it is predicted that Hungarian listeners will have an advantage over Polish listeners in word-final voicing perception. The two hypotheses outlined above, however, diverge as to how that advantage will affect performance in each of the two contexts. The L1 phonetic interference hypothesis would predict that Hungarian listeners should perform equally as well in both contexts, since the same exact acoustic cues are available. Perception is predicted to be good in both contexts, due to Hungarian listeners' L1 experience attending to word-final voicing cues. The L1 phonological interference hypothesis would predict that Hungarian listeners should perform less accurately when word-final consonants are in pre-obstruent position, since the L1 regressive assimilation rule will interfere only in pre-obstruent position.

As for Polish listeners, the L1 phonetic interference hypothesis predicts that they should do poorly in both contexts, due to the fact that they do not have much L1 experience attending to word-final voicing cues. The L1 phonological interference hypothesis makes the same prediction, but for different reasons. Under this hypothesis, Polish listeners should do poorly in the pre-pausal condition due to interference of their L1 final devoicing rule, and in the pre-obstruent condition due to interference of their L1 regressive voicing assimilation rule.

With respect to the different voicing contexts (voiced-voiced, voiced-voiceless, voiceless-voiced, voiceless-voiceless), the two hypotheses make different predictions. The L1 phonetic interference hypothesis predicts that performance across contexts in the pre-obstruent condition should reflect performance in the pre-pausal condition, and should reflect the strength of the voicing cues available. The L1 phonological interference hypothesis would predict that all four voicing contexts should be equally difficult since, consistent with regressive voicing assimilation, listeners should expect the voicing identity of the word-final obstruent to be obscured by the voicing of the following word-initial consonant. That is, in a context with matched voicing, such as in the name *Jess Keller*, non-native listeners will not know if the word-final /s/ is an assimilated /z/ or an underlying /s/. Similarly, in a context with mismatched voicing like *Jess Geller*, non-native listeners may interpret the /s/ as a /z/, since /z/ would be consistent with the voiced initial /g/.

<u>Question 3</u>: How will Hungarian and Polish listeners perceive English word-initial voicing contrasts?

Predictions:

The L1 phonetic interference hypothesis would predict that perception of word-initial voiceless stops should be good, since the long-lag VOT values of English voiceless stops corresponds to the category boundary of voiceless stops for both Hungarian and Polish. Perception of voiced stops should depend on whether or not they were produced with negative or short-lag VOT. Since voiced stops in voiced-voiced contexts will likely be produced with negative VOT, perception is predicted to be good. Voiced stops in voiceless-voiced contexts may be produced with short-lag VOT, so in this context they may be confused with voiceless stops. Perception of fricatives should be good, as long as they are produced with sufficient voicing cues. Again, as with word-final perception, accuracy rates in both post-obstruent and post-pausal conditions should be the same.

The L1 phonological interference hypothesis would predict that word-initial obstruents will be more successfully perceived in sequences with matched voicing (e.g. voiced-voiced or voiceless-voiceless) than mismatched voicing. In a stimulus such as *Jess Geller*, the voicelessness of the first obstruent may create an expectation that the second obstruent will also be voiceless. While an extreme version of this hypothesis would predict that word-initial segments both voiced-voiceless and voiceless-voiced sequences should be equally difficult, it is likely that the strength of the cue to voicelessness in word-initial stops (e.g. long-lag VOT) might override any anticipatory effects from the voicing of the word-final obstruent.

Chapter 2. Method

2.1. Participants

Participants were 12 native Polish (PS) speakers, 11 native Hungarian (HU) speakers, and 8 native American English (AE) controls. The native Polish and Hungarian speakers were all living in the US at the time of testing, and all had arrived in the US after the age of 18 years old. All participants passed a hearing screening (ANSI standards 25 dB HL at 500, 1000, 2000 and 4000 Hz). Information about the participants' language background, such as length of residence (LOR) in English-speaking countries, formal education in English, and daily English usage was collected by means of a questionnaire. Language questionnaires can be found in Appendices A and B. Participants received a small stipend for their participation.

The AE participants were native speakers of American English and had all grown up speaking English only, with parents who were fluent in English. They ranged in age from 21-46 years old. The HU participants ranged in age from 25-43 years (Median = 30 years), and had lived in the US (or other English speaking country) for a range of 11 months to 11 years (Median = 6.5 years). The PS participants ranged in age from 23-34 (Median = 30.5) and had lived in the US for a range of 7 months to 8 years (Median = 4.75). None of the non-native participants (with one exception) were fluent in languages other than English and their native language. One Polish participant was fluent in Russian, but was included in the study since Russian, like Polish and Hungarian, has a regressive voicing assimilation rule.

An additional 10 participants (3 AE, 4 HU, 3 PS) took part in the study, but their data were not included for various reasons. Data from three AE participants were not used in the final analysis due to the fact that their parents were not born and raised in the United States. Data from three Hungarians were not included in the final analysis since they reported being fluent in German. (German, like English, has voiceless aspirated stops). Data from an additional Hungarian was excluded because he reported having difficulty reading in his native language. Data from two Polish speakers were excluded because they arrived in the United States before the age of 18 years old. Finally, data from one Polish participant was excluded because she failed the hearing screening.

The duration of the experiment was 1½ to 2 hours. After the hearing screening, the perception task was administered, followed by the production task and language background questionnaire.

2.2. Perception Task

2.2.1. Stimulus materials

The stimuli for the perception study were first and last name combinations, embedded in the carrier phrase *'I met* _________ *today'*. The stimuli were constructed so that there would be three different manner combinations for the obstruent sequences: stop-stop, fricative-stop, and stop-fricative. For the first names, there were two sets of minimal pairs ending in fricatives (/s/ - /z/), two sets ending in labial stops (/p/ - /b/), and 2 sets ending in velar stops (/k/ - /g/). For the last names, there were 2 sets of minimal pairs beginning with stops (one labial, /b/ - /p/, and one velar, /g/ - /k/), and 2 sets of minimal pairs beginning with fricatives (/s/ - /z/). The stimuli are summarized in Table 2.1.

Each of the 24 full names was produced a total of 6 times, for a total of 144 tokens. All 144 tokens were used in the perception experiment. In addition to the 24 target items, an additional 24 filler items were created. These items were also minimal quadruplets, but do not contain the target obstruent sequence (e.g. *Ben/Belle Henley/Fenley)*. Each was produced four times, for a total of 96 filler items. Of these 96 fillers, a total of 16 tokens were used as part of the practice portion of the perception task, described in the next section. The full list of target stimuli and practice items, and number of repetitions each can be found in Appendix C.

A female native American English speaker from the New York area recorded the stimuli. The stimuli were randomized and presented in a recording script (sample protocol page in Appendix D**)**. The recording script consisted of 16 pages with 19 sentences per page. The first and last sentences of each page of the script were always the same; the final productions were not used in the experiment to eliminate issues with list-final intonation. The speaker was instructed to speak naturally, yet "slow enough so that a listener would be able to discern which name she was saying". This instruction was included to minimize listener error. The stimuli were normalized to -20 dB. Two native English speakers with backgrounds in phonetics verified the stimuli for intelligibility.

The tokens were modified using waveform editing procedures so that, for each token, there was a version with the last name removed (e.g. *I met Jess [silence] today*) and a version with the first name removed (e.g. *I met [silence] Geller today*), in addition to the full version, for a total of 432 tokens. Cuts were made as follows:

Fricative-stop: In the closure after the offset of high frequency frication, before the onset of any prevoicing

Stop-stop: In the closure after the release of the first stop, before the onset of any prevoicing

Stop-fricative: Since the burst of the stop was, in most cases, inseparable from the fricative, cuts were made in the closure prior to the release of the first stop.

In cases where the cut occurred in the middle of phonation (rather than silence), the cuts were potentially noticeable. To minimize effects of the unnaturalness of the cut, 10 millisecond onset/offset ramps were imposed to certain stimuli: word-final voiced [b, z] before a voiced obstruent, and word-initial [b, g] after a voiced obstruent. For word-initial [s, z] in all environments, a 40 millisecond amplitude onset ramp was imposed.

2.2.2. Procedures

The perception experiment was comprised of two tasks: a two-choice identification task and a four-choice identification task. The two-choice identification task contained the edited first and last name tokens (e.g. *I met Jess [silence] today*, with response choices *Jess* and *Jez*). The four-choice identification task contained the full name tokens (e.g. *I met Jess Geller today*, with response choices *Jess Geller, Jez Keller, Jess Geller, Jess Keller*). All stimuli were randomized and presented using a customized, computer-controlled experimental procedure (Paradigm ID by Tagliaferri, 2007). Each stimulus was presented over headphones (MAICO TDH 39). After the stimulus was completed, the response alternatives appeared on the computer screen. Participants responded by clicking on one of the alternatives, after which the experiment continued.

The 288 edited tokens for the two-choice task were randomized into 6 blocks of 48 sentences each. Both the first- and last-name edited tokens were included within the same blocks, so that each block had equal numbers of first- and last-name tokens. The 144 full name tokens for the four-choice task were randomized into 3 blocks of 48 sentences each. All together, there were 9 blocks of 48 sentences each. Blocks 3, 6, and 9 contained the full name stimuli (four-choice task), and the rest of the blocks contained the edited stimuli (two-choice task). A break was given after completion of 6 blocks. Participants were informed that they could take a break at any time before or after the scheduled break.

Prior to presentation of the experimental trials, participants read instructions on a computer screen and completed a brief practice session. See Appendix E for the full text of the instructions. The practice session consisted of non-experimental stimuli (*Ben/Belle Henley/Fenley*), and was presented in two blocks of 16 tokens each. The first practice block was a two-choice task with edited stimuli (e.g. *I met Ben [silence] today; I met [silence] Henley today*). The second practice block was a four-choice task with full name stimuli (e.g. *I met Ben Henley today*). The experimenter stayed with the participant during the practice session and answered any questions at that time. Once the participants understood the task, the experimental trials began.

2.2.3. Acoustic analysis of stimuli

Acoustical analysis of the stimuli was performed in Praat (Boersma and Weenink, 2007). Durations were measured as follows.

<u>Fricative-Stop Sequences:</u>

Figure 2.1 shows a sample waveform and spectrogram illustrating the measurement criteria for the fricative-stop sequences.

Preceding vowel duration: Measured from the onset of upper formant periodic energy to the beginning of the fricative.

Fricative Duration: Measured from the onset to the offset of high-frequency frication (4000 – 8000 Hz).

Fricative Voicing Duration: Voicing was measured based on evidence of regularly spaced vertical striations at the level of the baseline on the spectrogram, and by evidence of periodicity on the waveform. The total fricative duration was divided by the fricative voicing duration (multiplied by 100) to obtain the percentage of voicing present during frication.

Closure Duration: Measured from the offset of intense, high-frequency frication (4000 – 8000 Hz) to the release burst of the following stop.

Stop voicing: Measured from the onset of regularly spaced vertical striations at the level of the baseline on the spectrogram to the onset of the release burst, if present (negative VOT) OR from the onset of the release burst to the onset of regularly spaced vertical striations (short- or long-lag VOT).

<u>Stop-stop Sequences:</u>

Figure 2.2 shows a sample waveform and spectrogram illustrating the measurement criteria for the stop-stop sequences.

Preceding vowel duration: Measured from the onset to the offset of upper formant periodic energy.

Closure and Voicing Durations: Stop-stop sequences were measured differently depending on whether a release burst was present. A release burst was counted if it was audible and strong enough to create a band of energy on the spectrogram from the level of the baseline to the higher frequencies. Closure and voicing durations were measured as follows:

If release burst for stop 1 present:

Closure 1 Duration: Measured from the offset of the preceding vowel to the release burst for stop 1.

Release burst + low amplitude noise Duration: Measured from the onset of the release burst (first deflection on the waveform after closure for stop 1 and abrupt onset of energy on spectrogram) to the end of noise visible on the spectrogram.

Closure 2 Duration: Measured from the end of the release burst for stop 1 to the beginning of the release burst for stop 2.

Stop 1 voicing: Measured during closure 1 as evidenced by regularly spaced vertical striations at the level of the baseline of the spectrogram.

Stop 2 voicing: Measured from the onset of regularly spaced vertical striations at the level of the baseline on the spectrogram to the onset of the release burst, if present (negative VOT) OR from the onset of the release burst to the onset of regularly spaced vertical striations (short- or long-lag VOT).

If no release burst for stop 1 present:

Closure 1 duration: Measured from the offset of the preceding vowel to the middle of the closure.

Closure 2 duration: Measured from the middle of the closure to the beginning of the release burst for stop 2.

Stop 1 voicing: Measured during closure 1 as evidenced by regularly spaced vertical striations at the level of the baseline of the spectrogram.

Stop 2 voicing: Measured from the onset of regularly spaced vertical striations at the level of the baseline on the spectrogram to the onset of the release burst, if present (negative VOT) OR from the onset of the release burst to the onset of regularly spaced vertical striations (short-lag or lead VOT).

Stop-Fricative Sequences:

Figure 2.3 shows a sample waveform and spectrogram illustrating the measurement criteria for the stop-fricative sequences.

Preceding vowel duration: Measured from the onset to the offset of upper formant periodic energy.

Closure Duration: Measured from end of preceding vowel to the release burst (first deflection on the waveform after the closure and abrupt onset of energy on spectrogram).

Voicing in closure: Voicing was measured based on evidence of regularly spaced vertical striations at the level of the baseline on the spectrogram, and by evidence of periodicity on the waveform.

Release Burst Duration + low amplitude noise: Measured from the onset of the release burst (first deflection on the waveform after closure for stop 1 and abrupt onset of energy on spectrogram) to the end of noise visible on the spectrogram.

Fricative Duration: Measured from the onset to the offset of high-frequency frication (4000 – 8000 Hz).

Fricative Voicing Duration: Voicing was measured based on evidence of regularly spaced vertical striations at the level of the baseline on the spectrogram, and by evidence of periodicity on the waveform. The total fricative duration was divided by the fricative voicing duration (and multiplied by 100) to obtain a percentage of voicing present during frication.

The median durations and ranges of the relevant acoustic segments for word-final consonants are shown in Table 2.2. Appendix F gives a more complete summary of the same data, broken down by each individual stimulus type. The percentage of voicing in a fricative or closure appeared to reliably distinguish voiced versus voiced obstruents in these tokens. For word-final fricatives, the percent fricative voicing was always larger for voiced (Median = 78.3%) than voiceless (Median = 8.1%) fricatives. Some devoicing did occur before voiceless stops (Medians = 39.8% and 37.3% for *Jez#K* and *Kaz#P*, respectively), but never to the level of the voiceless stops. Similarly, for word-final stops, voiced stops were almost always produced with 100% closure voicing, while voiceless stops were always produced with less. Full closure voicing for voiced stops was maintained even when preceding voiceless consonants.

Fricative duration also appeared to be a cue for voicing of word-final fricatives, with longer durations for voiceless (Median = 78 ms) versus voiced (Median = 47 ms) fricatives. The distributions for fricative durations were non-overlapping for minimal pairs.

Vowel durations tended to be shorter before voiceless final consonants in these tokens. This was very clear for the vowel /ɛ/ across all contexts, with near non-overlapping distributions for voiced versus voiceless word-final consonants. It was less clear, however, when the preceding vowel was /æ/.

For word-initial stops, VOT appeared to serve as a cue for voicing in these tokens. Table 2.3 summarizes the median VOT values and ranges for word-initial stops. Word-initial /b/ was almost always produced with negative VOT. The exception was in the context of *Mack Billman*, where it followed a voiceless velar stop. Here it was produced with a median VOT of 8 milliseconds, characteristic of short-lag VOT. The /p/ and /b/ distributions were non-overlapping, with VOT values for /p/ in the long-lag range (Median = 49 ms).

Word-initial /g/ was always produced with negative VOT after a voiced consonant (Median = -48 ms). With the exception of one token, it was produced with short-lag VOT following a voiceless consonant (Median = 12 ms). Word-initial /k/ was produced with long-lag VOT (Median = 60 ms). As with word-initial /p/ versus /b/, the VOT distributions for word-initial /k/ versus /g/ were non-overlapping.

For word-initial fricatives, the percentage of voicing during frication appeared to be a reliable cue for voicing in these tokens. Table 2.4 summarizes the median durations and ranges for word-initial fricatives. Word-initial /s/ was always produced with little (15% or less) or no voicing. Word-initial /z/ was frequently fully voiced, with median voicing at 100% in all contexts except *Mack Zimmer*. In the context of *Mack Zimmer*, two tokens of /z/ were devoiced to 12.5% and 15% voicing. These two instances of progressive devoicing will be taken into consideration when interpreting the perception results.

Fricative durations tended to be shorter for word-initial voiced /z/ (Median = 85 ms) than voiceless /s/ (Median = 120 ms). There was, however, some overlap in the distributions, so this did not appear to be a reliable cue for word-initial fricative voicing.

2.3. Production task

2.3.1. Stimulus materials and procedure

The production stimuli were created to mirror the perception stimuli: first and last name combinations embedded in the carrier phrase *I met* _____ *today*. Just as in the perception task, there were three different manner combinations for the obstruent sequences: stop-stop, fricative-stop, and stop-fricative. Near minimal pairs were used instead of minimal pairs, since it was felt this would make the task more manageable for non-natives. For the first names, there were two sets of near minimal pairs ending in fricatives (/s/ - /z/), two sets ending in labial stops (/p/ - /b/), and 2 sets ending in velar stops (/k/ - /g/). For the last names, there were 2 sets of near minimal pairs beginning with stops (one labial, /b/ - /p/, and one velar, /g/ - /k/), and 2 sets of minimal pairs beginning with fricatives (/s/ - /z/). See Appendix G to see the full list of stimuli for the production task.

Thirty additional fillers were created to break up the monotony of the productions for the speakers. These fillers contained the target contrast, but used a noun + verb combination instead of a full name, e.g. *The pu**p g**ives her a big kiss*). The fillers were randomized along with the target sentences and broken down into 4 lists for the participants to read. Before recording each list, the participants were instructed to read through each sentence aloud to familiarize themselves with the words. Once the participants were comfortable with each list and given the opportunity to ask the experimenter questions, the recording session began. Participants were instructed to read each sentence twice during the recording session, since it was discovered during piloting that this helped speakers to get into a rhythm. After completion of each list, the experimenter asked the speaker to repeat any items that needed repeating (due to disfluencies such as an extra long pause or stumble, or interference from an external noise).

Chapter 3. Results

In the experiment, the perception task preceded the production task. The results of the production task are reported here first, however, in order to provide perspective for the perception results.

3.1. Production task

3.1.1. Data analysis

In the production task, participants read a list containing two repetitions each of 24 target stimuli, resulting in a total of 48 tokens for acoustic analysis. Participants were instructed to read each sentence in the protocol at least twice, since it was discovered during pilot testing that this helped participants feel more comfortable with each utterance. (Further details on the procedure are specified in Chapter 2).

In selecting tokens for analysis, the second of the two repetitions was always chosen unless there was a disfluency in that utterance; in those cases, the first utterance was retained for analysis. Once the 48 tokens were selected for each of the participants, acoustic analysis was performed in Praat (Boersma and Weenink, 2007). The criteria used for the acoustic analysis were the same that were used for the perception stimuli, detailed in Chapter 2.

Since two tokens of each stimulus were analyzed per speaker, an issue to be addressed was whether the measurements of these two tokens should be averaged to create one single score. It was decided that this would not be feasible, since any two utterances produced by any one participant could potentially be categorically different. For example, if a speaker produced the /b/ in *Gus Barker* with -50 milliseconds of VOT in one instance, and +10 milliseconds of VOT in the next, it would be undesirable to average these two scores to -30. Averaging the scores would obscure the fact that one instance had negative VOT, while the other had short-lag VOT.

In order to avoid this type of misrepresentation of the data, it was decided that each of a speaker's two tokens would be analyzed together with the rest of the data from that speaker's language group. The American English group, for example, has 16 tokens (8 speakers X 2 tokens) for any given acoustic cue.

Tokens were measured according to the criteria outlined in Chapter 2 by the experimenter. To verify consistency of the acoustic measurements, 10% of the tokens were measured a second time. The average difference between the first and second measurements for the relevant variables is reported in Appendix I.

A summary of the acoustic measurements across language groups is listed in Appendix J. Tables 1-3 show medians and inter-quartile ranges for relevant cues in the fricative-stop sequences: vowel duration, fricative duration, percent fricative voicing, and VOT. As mentioned above, speakers employed negative versus positive VOT at variable rates. While all voiceless stops had VOT in the positive range, voiced stops had both negative and positive VOT. The data for VOT is divided so that positive and negative VOT are listed separately.

Tables 4-6 show medians and inter-quartile ranges for relevant cues in the stop-stop sequences: vowel duration, percent voicing duration, and VOT. The stop-stop sequences could be produced either with or without a word-final stop burst. These instances are separated out in the table.

Tables 7-9 show medians and inter-quartile ranges for relevant cues in the stop-fricative sequences: vowel duration, percent closure voicing, fricative duration, and percent fricative voicing.

The following sections explore the data from the acoustic analysis in more detail. Sections 3.1.2. – 3.1.6. discuss statistical differences in word final voicing and duration. In these sections, the issue is the extent to which voicing of one segment changes as a function of the voicing of the adjacent segment. Section 3.1.7. discusses observations with respect to vowel duration and VOT.

3.1.2. Method of analysis – non-parametric statistics

Since the amount of data for each individual cue analyzed in this sample was small and not normally distributed, non-parametric statistics were used for the analysis. To compare any two groups on a single variable, the median score between the two groups was determined. The number of scores from each group above or below the median was then calculated. Scores that were equal to the median were included in the 'above' category, unless the median happened to be zero, in which case they were counted in the 'below' category. These scores were then tested for significance using a Fisher's Exact test. For each statistical test that was done, a corresponding table lists the relevant median scores and the significance levels.

3.1.3. Word-final fricative voicing and duration

Word-final fricative voicing was analyzed as a percentage (voicing duration during frication divided by the fricative duration). It was predicted that native American English speakers would maintain a voicing distinction word-finally, but that the non-native groups would regressively voice or devoice word-finally based on the voicing of the following obstruent. The target stimuli were *Buzz/Gus Barker/Potter* and *Liz/Siss Guffman/Kenner*.

Results for the voicing percentages for the three language groups are summarized in Table 3.1. Word-final fricative durations are summarized in Table 3.2. For the AE group, /s/ remained voiceless regardless of voicing of the following obstruent. There were no significant differences in the percent voicing duration of word-final /s/ preceding voiced versus voiceless stops ($p > .05$). There was, however, a contextual effect for word-final /z/. Before voiceless stops, /z/ was often partially devoiced by native AE speakers. The median percent voicing for the /z/ in both *Buzz* and *Liz* was significantly lower in voiceless versus voiced stops ($p < .01$). Looking at fricative duration, only one contextual effect was observed. Word-final /s/ in *Siss* was significantly longer before voiceless than voiced stops ($p < .05$). This is interesting because there was no corresponding significant effect in voicing duration in this same context.

Native Hungarian speakers had a similar pattern with respect to devoicing of word-final /z/. The median percent voicing for the /z/ in both *Buzz* and *Liz* was significantly lower before voiceless versus voiced stops ($p < .001$). Word-final /s/ also showed a context effect. The median percent voicing for the /s/ in both *Siss* and *Gus* was significantly higher before voiced versus voiceless stops ($p < .01$ and $p < .05$, respectively). Corresponding contextual effects on fricative length were observed for *Liz/Siss*, with longer fricative durations before voiceless stops ($p < .001$). The same fricative duration effect was not found in *Buzz/Gus*.

The PS group patterned in almost exactly the same way as the HU group. Word-final /z/ was significantly more devoiced before voiceless stops. The median percent voicing for the /z/ in both *Buzz* and *Liz* was significantly lower in voiceless versus voiced stops ($p < .001$). The median percent voicing for the /s/ in both *Siss* and *Gus* was significantly higher before voiced versus voiceless stops ($p < .01$). Corresponding contextual effects on fricative length were observed for *Liz/Siss*, with longer fricative durations before voiceless stops ($p < 0.01$). The same fricative duration effect was not found in *Buzz/Gus*.

Comparing the non-native groups to the native American English group, both non-native groups showed more devoicing and regressive voicing assimilation than the native speakers. The Hungarian group devoiced /z/ significantly more than the American English group in all contexts ($p < .01$), except in the name *Liz Guffman.* This is due to the fact that the native American English speakers tended to devoice /z/ in this context, with 46% median voicing. Production of /s/ by the HU group did not significantly differ from the native AE group. The difference in voicing between word-final /s/ in AE versus HU *Gus Potter* barely reached significance ($p = .05$), but because the median percent voicing for both groups was so low (5.1%), this does not seem to be an exception to the rest of the findings.

Native Polish speakers devoiced /z/ significantly more than native American English speakers in all contexts ($p < .01$) except *Liz Guffman*. The difference between the AE and PS groups only just reached significance ($p = .05$) in the context *Liz Guffman*, which again was probably due to the fact that the AE speakers tended to partially devoice the /z/ in this context. Word-final /s/ voicing did not differ significantly between AE and PS groups in any context except *Siss Guffman,* where the PS group significantly voiced /s/ more than the AE group ($p = .01$)

In summary, all groups tended to show regressive devoicing of /z/, but the non-natives did so more than the native American English speakers. The non-natives tended to have more devoicing of /z/ in general, regardless of context. Native AE speakers did not show any significant regressive voicing of /s/ before voiced stops. In contrast, both native Hungarian and Polish speakers had more voicing in /s/ before voiced than voiceless stops. This was a slight tendency, however, and not robust enough to show a statistical difference from native AE speakers. The one exception was for the Polish speakers, who differed significantly from the native AE speakers in the context *Siss Guffman*.

3.1.4. Word-final voicing – stops (stop-stop context)

Word-final stop voicing was calculated by measuring the percentage of voicing present in the closure. The target stimuli were *Jug/Buck Barker/Potter* and *Gib/Kip Guffman/Kenner*. Analysis of the closure voicing in the stop-stop sequences was complicated by the fact that some tokens had released final stops, while others did not. This is illustrated in figure 3.1, which shows two separate tokens of *Jug Potter*: one with a released /g/, and one with an unreleased /g/. When the stop is released, the percentage of voicing in the closure prior to the release can be used as a measure of stop voicing. When a stop is unreleased, however, there is no clear endpoint for the final stop, and only one closure between the offset of the vowel and the release of the word-initial stop. In these cases percent closure voicing is misleading as a measure of stop voicing. In the case illustrated in Figure 3.1, the closure is 100% voiced in the token with the release, but 55% voiced in the token without the release. It is likely that in both cases, the /g/ could be considered fully voiced, but only the first token is reflective of this. In comparing final stop voicing, then, an adjustment was made. For those tokens with unreleased stops, if the amount of voicing in the closure was more than 0% and less than 100%, the percent closure voicing was multiplied by two (and capped at 100%). So in the case illustrated in Figure 3.1, the 55% single closure voicing would be multiplied to be 110%, then reduced to 100%. This adjustment was necessary on less than 10% of the stop-stop tokens.

Results for the three language groups are summarized in Table 3.1. Voicing in word-final stops did not significantly differ by context for the AE group. Median percent voicing for word-final voiced stops was 100% regardless of context. Word-final /k/ had median percent voicing of 0% regardless of context. Word-final /p/ always had a certain amount of voicing, but there was no significant difference between the two voicing contexts.

In contrast, the HU group did show context effects on word-final stop voicing. Word-final stops were always voiced less when they preceded voiceless stops. Word-final voiced stops /b/ and /g/ had median percent voicing of 100% before voiced stops, and significantly less before voiceless stops ($p < .001$). Similarly, word-final voiceless stops in *Buck* and *Kip* became more voiced when they preceded voiced stops ($p < .001$, $p < .05$, respectively). This pattern of results suggests that the Hungarian group tended to assimilate the voicing of word-final stops to the voicing of the following stop.

The Polish group showed context effects for all word-final stops except /p/. Word-final voiced stops /b/ and /g/ both had significantly greater voicing before voiced than voiceless stops, ($p < .001$). For word-final voiceless stops, word-final /k/ had significantly more voicing before voiced stops ($p < .01$). Again, like the native

Hungarian speakers, the native Polish speakers appeared to assimilate word-final voicing to the voicing of the following stop.

Comparing the AE group to the two non-native groups, there were significant differences in the patterns of voicing. Before voiced stops, there was no significant difference between the AE and HU groups in word-final /g/, but there was a significant different for word-final /b/ ($p < .01$). Given that the median percent word-final /b/ voicing was 100%, the difference here seems one of degree rather than a categorical difference – the HU group showed more devoicing than the AE group in this context, but as a whole, word-final stops remained voiced before voiced stops. Word-final voiced stops, however, were significantly more devoiced before voiceless stops by the HU group ($p < .001$). The HU group also showed more evidence of regressive voicing than the AE group in one context. Before a voiced stop, the HU group had a higher median voicing percentage for voiceless /k/ than the AE group ($p < .01$).

Comparisons between the AE and PS groups showed a similar pattern of results. Before voiced stops, there was no significant difference between the two groups in word-final /g/, but there was a significant difference for word-final /b/ ($p < .01$). Again, the median percent word-final /b/ voicing was 100% for both groups, so any observed tendency of the PS group to devoice in this context was not very pronounced. Before voiceless stops, in contrast, word-final /b/ and /g/ were significantly more devoiced by the PS group ($p < .001$). The PS group also showed more evidence of regressive voicing than the AE group. Before voiced stops, both /k/ and /p/ were voiced significantly more ($p = .01$ and $p = .05$, respectively) by the PS group than the AE group.

In summary, in the stop-stop context, the native American English speakers did not show any evidence of regressive voicing assimilation. Word-final stops retained their underlying voicing characteristics regardless of the voicing of the following stop. The non-native groups, however, did show evidence of regressive voicing assimilation in this context. Both the Hungarian and Polish groups devoiced word-final /b/ and /g/ before voiceless stops. Both groups had a slight tendency to voice word-final /p/ and /k/ before voiced stops, but the tendency was stronger in the Polish group than the Hungarian group.

3.1.5. Word-final voicing – stops (stop-fricative context)

Word-final stops behaved somewhat similarly before fricatives to those before stops, but there were differences. Results for the three language groups are shown in Tables 3.1. The target stimuli were *Jug/Buck Zelder/Simpson* and *Gib/Kip Zelder/Simpson*.

The AE group had consistent word-final stop voicing regardless of the voicing of the following fricative. The one possible exception was word-final /b/, which was devoiced significantly more before /s/ than /z/ ($p = .05$). This result, however, was based on the fact that 4 tokens of word-final /b/ were voiced less than 100%, but considering the least voiced of these 4 tokens had 75% voicing in the closure, it still seems fair to say that the native AE group maintained voicing even before voiceless fricatives.

The native HU group, in contrast, had significant differences in word-final voicing based on the voicing of the fricative that followed. Word-final voiced stops had significantly greater amounts of voicing before voiced than voiceless fricatives ($p < .001$). Similarly, voiceless word-final stops were voiced more preceding voiced fricatives ($p < .001$). Like the HU group, the PS group showed evidence of regressive devoicing before /s/. The median percent voicing of voiced stops was significantly higher preceding voiced than voiceless fricatives ($p < .001$). The PS group did not, however, appear to regressively voice stops preceding /z/. The differences between word-final voiceless stops in the two voicing contexts were not significant.

Comparing the AE and HU groups, both groups had median word-final voicing of voiced stops at 100% before /z/. There was a difference between the AE and HU groups that reached significance – final /b/ was devoiced more often by the HU than AE group ($p = .05$). But this appears to be a result of the strictness of the statistical test. Only 5 of 22 HU tokens were voiced less than 100%, 4 of which were voiced above 60%. So it seems fair to say that the HU group maintained final stop voicing in this context despite the significant result. Before voiceless /s/, however, the HU devoiced word-final voiced stops to a much greater degree than the AE group ($p < .001$). Compared to the AE group, the HU group did not regressively voice voiceless stops to a significantly greater degree.

A comparison of the AE and PS group shows that the PS group had a tendency to devoice word-final voiced stops across all contexts. Before /z/, while many of the native Polish speakers maintained closure voicing through 100% of the closure, some did devoice. The difference between the PS and AE groups in this context was significant ($p < .05$ for word-final /g/, $p < .01$ for word-final /b/). Devoicing by the PS group was even more pronounced before voiceless /s/. In this context, the difference between the two groups was more robust ($p < .001$). There was no significant difference, however, between the two language groups for word-final voiceless stops. The native PS group did not regressively voice any more than the AE group.

In summary, while the AE group maintained the same degree of word-final voicing regardless of context, both the PS and HU groups showed signs of regressive voicing assimilation in the stop-fricative context. Both the HU and PS groups regressively devoiced word-final /b/ and /g/ before /s/. The PS group, but not the HU group, additionally showed some devoicing of stops before voiced /z/ as well. The HU group, but not the PS group, showed signs of regressively voicing stops preceding /z/. This tendency, however, was not robust enough to show a significant difference between the HU and AE groups in this context.

3.1.6. Word-initial fricative voicing and duration

Word-initial fricative voicing was analyzed as a percentage (voicing duration during frication divided by the fricative duration). Whether the fricative voicing began at the beginning or middle of the fricative was not differentiated for this analysis. Results for word-initial fricative voicing and duration by the three language groups are shown in Tables 3.3 and 3.4. Word-initial /s/ was voiceless across all language groups in all contexts, so there is little to discuss here. The one possible exception is the voicing in /s/ in *Gib Simpson* produced by the American English speakers. A few tokens were produced with between 6% - 12.5% voicing. This is a small percentage, but a noticeable trend since the HU and PS groups produced 0% voicing in word-initial /s/ across the board. There was also one AE token with 76% 'voicing', but this token appears to be an anomaly. It sounds to the experimenter's ears like an /s/, but there is some periodicity throughout the frication that is perhaps due to a cold or some other articulatory inconsistency.

The AE group produced word-initial /z/ with median 100% voicing after all stops except /k/. The difference in /z/ voicing following /k/ versus /g/ was significant ($p = .01$). This context effect was also reflected in fricative duration – with longer median /z/ durations following /k/ versus /g/ ($p < .05$). No context effects were found following labial stops.

The Hungarian group tended to devoice word-initial /z/, with median percent voicing across all contexts less than 80%. While there was a tendency for fricatives to have greater amounts of voicing following voiced stops, this tendency was not significant. There was, however, a context effect for fricative length, with longer median /z/ durations following /k/ versus /g/ ($p < .05$).

The Polish group maintained voicing in word-initial /z/ regardless of context. Median voicing was 100% following both velars and labials, and the difference in voicing following voiced versus voiceless stops was not significant. Accordingly, there were no observed context effects on fricative duration.

Comparing the AE and HU groups, word-initial /z/ was voiced significantly more by the AE group than the HU group following labial stops ($p = .001$) and /g/ ($p = .05$). The difference in /z/ voicing between the two groups following /k/ was not significant, due to the fact that this was the one context where the AE group tended to devoice /z/. The AE and PS groups did not differ significantly in word-initial /z/ voicing in any context.

In summary, it appears that in this sample the PS group had a tendency to fully voice /z/ regardless of context, while the HU group had a tendency to devoice /z/

regardless of context. The AE group tended to maintain /z/ voicing in most contexts, but did show evidence of progressively devoicing /z/ after /k/.

3.1.7 Other cues

Vowel duration was also measured, and is another potential cue to final obstruent voicing. Median vowel durations and inter-quartile ranges are displayed in Table 3.5. (Full summary data for each utterance type are reported in Appendix J, Tables 1 – 9). Since it is difficult to control for speaking rate when looking at vowel durations within and across groups, statistical tests were not performed, but some generalizations can be made. For the AE group, median vowel durations were longer before voiced obstruents than before their voiceless counterparts. The inter-quartile ranges for vowel durations before voiced versus voiceless obstruents were non-overlapping, suggesting that the tendency was consistent. While the non-native groups showed the same general trend as the native American English speakers, with lower median vowel durations before voiceless obstruents, the distinction was not as clear. Inter-quartile ranges for non-native vowel durations before voiced versus voiceless obstruents tended to have quite a bit of overlap.

Word-initial VOT was also measured, and median values and ranges are reported in Table 3.6. (Full summary data is displayed in Appendix J). For initial voiced stops, all groups alternated between producing short-lag and negative VOT. The AE group employed short-lag the most out of the three groups, followed by the HU group, then the PS group. All groups prevoiced initial /b/ more than initial /g/. For voiceless word-initial stops, all groups produced initial /p/ and /k/ with longer VOT values than short-lag /b/ and /g/. The non-native groups, however, had shorter VOT values than the AE group.

3.2. Perception task

3.2.1. Data analysis

The first step in the perception data analysis was to compare the results of the two-choice versus four-choice task. The purpose of this first analysis was to determine how a given segment was perceived with and without the presence of an adjacent obstruent. The two-choice task, as will be recalled from the previous chapter, consisted of edited tokens with either the first- or last-name removed (e.g. *I met [silence] Geller today; I met Jess [silence] today*). In the four-choice task, the full name was present (e.g. *I met Jess Geller today*). To compare the two contexts, the percent accuracy scores were analyzed for the word-final and word-initial segments separately. It should be noted that for the four-choice task, two of the four response alternatives had the same voicing value. That is, if the answer to *'I met Jess Geller today'* was 'Jess Geller', participants were scored as perceiving word-final /s/ correctly if they marked 'Jess Geller' or 'Jess Keller'.

The second step in the perception data analysis was to determine how different voicing contexts (e.g. voiced-voiced, voiced-voiceless), affected perception. Sections 3.2.2. and 3.2.3. discuss word-final perception in the two- versus four-choice tasks. Section 3.2.4. discusses the effect of the different voicing contexts on word-final perception. Section 3.2.5. discusses word-initial stop perception, and section 3.2.6. discusses word-initial fricative perception.

3.2.2. Word-final perception in the two- and four-choice tasks: AE versus HU & PS groups

Initial checks for homogeneity of variance revealed that the error variances across groups were not equal. For the two-choice task, Levene's test was significant for the stop-stop variable, $F(2, 28) = 3.937$, $p < .05$. For the four-choice task, Levene's test was significant for all variables (fricative-stop: $F(2, 28) = 3.215$ $p = .055$; stop-stop: $F(2, 28) = 4.356$, $p < .05$; stop-fricative: $F(2, 28) = 3.521$, $p < .05$). The heterogeneity of variance was due to the fact that the native American English control (AE) group was performing at ceiling. Because of this, non-parametric statistics were used to compare native versus non-native performance.

The median word final accuracy scores for the two-choice task are displayed in Table 3.7, and the scores for the four-choice task are displayed in Table 3.8. The scores for the AE group were highly accurate, with group median accuracy scores at or above 90% across all conditions in the two-choice and four-choice tasks. The greater cognitive demands of the four-choice task did not appear to affect accuracy for this group, with accuracy even improving in the four-choice task in the stop-stop and stop-fricative contexts. The individual accuracy rates mirror the group results – all participants had median accuracy scores at or above 90% overall in both tasks. Individuals in the AE groups were also internally consistent -- all participants had median accuracy scores of at least 90% in 2 out of the 3 manner conditions.

The scores for the non-native groups were not as accurate as the AE group. It should be noted, however, that some individual participants were more accurate than others. In the HU group, 4 out of 11 participants scored 90% or higher overall in the two-choice task, and 5 out of 11 did so in the four-choice task (3 of these were the same individuals). In the PS group, 5 out of 12 participants scored 90% or higher overall in the two-choice task, but only 1 out of 12 did so in the four-choice task. There was, then, a small portion of each group that had consistently high scores.

Figure 3.2 illustrates the overall percent accuracy in the two-choice task across the three language groups. Using a Mann-Whitney U test to compare the language groups, the non-native groups were lower in overall percent accuracy than the AE controls in the two-choice task. The difference between the HU and AE groups was significant (U = 15.0, $p < .01$). The difference between the PS and AE groups was also significant (U = 11.5, $p < .01$). All p values are derived from one-tailed tests.

When scores in each manner condition were compared, the HU group remained significantly poorer than the AE group in the fricative-stop context (U = 24.0, $p < .05$) and stop-fricative context (U = 16.5, $p = .01$), but not the stop-stop context (U = 31.5, $p > .05$). Performance by the PS group was significantly poorer than the AE group in all contexts (fricative-stop: U = 22.5, $p < .05$, stop-stop: U = 26.5, $p < .05$, stop-fricative: U = 14.0, $p < .01$).

Figure 3.3 illustrates the overall percent accuracy in the four-choice task across the three language groups. On the four-choice task, the HU group performed significantly more poorly overall than the AE group (U = 8.0, $p < .01$). Comparison of the overall accuracy scores of the AE and PS groups was also highly significant (U = 1.0, $p < .001$), showing poorer overall performance by the PS group.

Breaking the same scores down into the different manner conditions, the HU group was significantly poorer than the AE group in the stop-stop context (U = 7.0, $p = .001$) and the stop-fricative context (U = 9.0, $p < .01$), but not the fricative-stop context (U = 29.0, $p > .05$). The PS group was significantly poorer from the AE group in all manner conditions (fricative-stop: U = 5.5, $p < .001$; stop-stop: U = 2.5, $p < .001$; stop-fricative: U = 2.0, p < .001). This suggests that the difference in the four-choice condition was more robust for the Polish group than the Hungarian group.

In summary, both non-native groups perceived word-final obstruents significantly less accurately than the American English controls in both tasks. This finding appears to be more robust with the Polish than the Hungarian group. While the Polish group performed significantly more poorly in all conditions in both tasks across the board, the Hungarian group did perform as accurately as the AE controls in two manner conditions. More detailed analysis of the performance of the Hungarian and Polish groups is discussed in the following section.

3.2.3. Word-final perception in the two- and four-choice tasks: HU versus PS groups

A mixed design ANOVA was performed to compare the performance of the HU and PS groups in the two-choice and four-choice tasks across the three manner conditions. The two within-subjects factors were Sequence Type (fricative-stop, stop-fricative, stop-stop) and Task (two- vs. four-choice). Language was a between-subjects factor. Homogeneity of variance was assumed, since Levene's test of error variances was not significant for any of the variables in the two-choice task (fricative-stop: F(1, 21) = .304; stop-stop: F(1, 21) = 1.666; stop-fricative: F(1, 21) = .580), or the four-choice task (fricative-stop: F(1, 21) = .545; stop-stop: F(1, 21) = .068; stop-fricative: F(1, 21) = 3.663).

The main effect of Task reached significance, F(1, 21) = 44.104, p < .001, but the main effects of Sequence Type, F(1, 21) = 2.513, and Language, F(1, 21) = 3.588, did not. There was also a significant interaction of Task and Language, F(1, 21) = 12.833, p < .01. All p values are based on two-tailed tests.

Further analysis indicated that both the Hungarian and Polish groups had lower accuracy scores in the four-choice task than the two-choice task. A t-test for correlated samples revealed that the HU group performed more accurately in the two-choice task (M = 87.4%) than the four-choice task (M = 83.3%), t(10) = 2.537, p < .05). The same effect was found for the PS group (Ms = 86.2% vs. 72.5%, t(11) = 6.544, p < .001) Comparing the two language groups, it was found that while the Hungarian and Polish groups did not differ significantly on the two-choice task, t(21) = .414, there was a significant difference between the two groups on the four-choice task, t(21) = 2.829, p = .01.

In summary, when word-final obstruents were presented before an inserted pause in a simple two-choice task, the HU and PS groups were significantly less accurate than the AE controls, but not significantly different from each other. When the word-final obstruents were presented in their full context in a four-choice task, both groups became even less accurate, with the Polish listeners showing a more severe effect than the Hungarian listeners.

3.2.4. Effect of voicing context on word-final perception

Table 3.9 shows the overall median word-final percent accuracy scores across the different voicing contexts (voiced-voiced, voiced-voiceless, voiceless-voiced, and voiceless-voiceless) for the four-choice task. To evaluate differences between scores of the non-native groups from the American English controls, non-parametric statistics were used. To compare the performance of the HU versus the AE groups, the median of the joint scores of the AE & HU groups was taken. It was then determined how many scores from each language group were above and below the median. The same analyses were run to compare the performance of the AE and PS groups. All p values were derived from one-tailed Fisher's exact tests.

Collapsing across all manner conditions, word-final perception accuracy for the HU group differed significantly from the accuracy of the AE group in the voiced-voiceless and voiceless-voiceless contexts ($p < .05$). Breaking down the results into the different manner conditions, the results do not remain as clear. Results are presented in Appendix H. In the fricative-stop condition, the HU and AE groups differed significantly only in the voiceless-voiced context ($p < .05$). In the stop-stop context, the HU and AE groups differed in both the voiceless-voiced and voiceless-voiceless contexts ($p < .05$). And in the stop-fricative context, the AE and HU groups differed in three contexts: voiced-voiceless, voiceless-voiced, and voiceless-voiceless. The clearest finding from these data is that, no matter how the data are broken down, the Hungarian group did not differ significantly from the American English controls in the voiced-voiced context.

Collapsing across all manner conditions, word-final perception accuracy for the Polish group differed significantly from the accuracy of the American English controls in all voicing contexts ($p < .05$ for voiceless-voiceless context, $p < .01$ for all other contexts). Breaking down the results into the different manner conditions, the same result was found in the stop-stop context, with the PS group differing significantly in all voicing contexts from the AE group ($p < .01$ for all voicing contexts). In the stop-fricative context, the PS group differed significantly from the AE group in three out of the four voicing contexts (voiced-voiceless, $p < .05$; voiceless-voiced, $p < .01$; voiceless-voiceless, $p < .05$). In the fricative-stop context, the PS group differed significantly from the AE groups in the voiced-voiceless context ($p < .01$), and the voiceless-voiced context ($p < .01$). It seems from this analysis that context effects were minimal for the PS group, but this may be due to the fact that there was a highly significant overall difference in accuracy between the AE and PS groups. The one effect that did emerge, however, was that the voiced-voiced context was the least different between the AE and PS groups. This is consistent with the results from the Hungarian group.

3.2.5 Effect of familiarity of first name

Given that some of the first names used in this study are unusual (for example, *Depp*), differences between first name minimal pairs were analyzed. The purpose of this analysis was to determine if less familiar names like *Depp* or *Jez* were more difficult to perceive. Results are summarized in Table 3.10. Wilcoxon signed ranks tests revealed that there was a significant difference in response accuracy between *Jess* and *Jez* in both the two-choice task ($Z = -2.306$, $p > .05$) and the four-choice task ($Z = -2.099$, $p > .05$), with *Jez* having higher accuracy scores than *Jess*. Since *Jez* is arguably the less common of the two, this result actually runs counter to the hypothesis that less common names are more difficult to perceive. The only other significant result was with the minimal pair *Mack* versus *Mag*, with *Mag* perceived less accurately in the four-choice task ($Z = -2.403$). Since *Mag* is the less common of the two names, it could be possible that participants tended to hear the more common alternative. Since this was the only contrast with a significant tendency in this direction, however, it appears that name familiarity did not play a signficant role in the word-final perception task. Note that even if name familiarity were a factor in the perception of *Mack* versus *Mag*, it could still be argued that the contrast itself was difficult; it is only the *direction* of the difficulty that would be called into question.

3.2.6. Word-initial stop perception

Word-initial stop perception in the two-choice task was highly accurate across all language groups. Results are shown in Table 3.11. All groups had median accuracy scores of 100% overall. In the four-choice task, accuracy scores remain fairly high, but the difference between the American English controls and both non-native groups was significant. Results are shown in Table 3.12. Mann-Whitney U tests were conducted since the error variances between groups were not equal. Again, p values were derived from one-tailed tests. Performance by the HU group was significantly poorer than the AE on overall initial stop perception ($U = 12.0$, $p < .01$). This finding was consistent in both the fricative-stop context ($U = 20.0$, $p < .05$) and the stop-stop context ($U = 15.5$, $p < .05$). Similarly, the PS group performed significantly more poorly than the AE group overall ($U = 7.5$, $p = .001$), and this finding was consistent in both the fricative-stop ($U = 12.0$, $p < .05$) and the stop-stop context ($U = 15.5$, $p < .01$).

It was predicted that word-initial voiced stops might be difficult for Hungarian and Polish listeners when preceded by a voiceless obstruent, since short-lag English /b, g/ are phonetically similar to Polish and Hungarian /p, k/ in this context. Both HU and PS groups were highly accurate in this context (Medians = 91.7% for both groups, collapsing across fricative-stop and stop-stop conditions). The HU group range of 66.7% - 100%, and the PS group range of 62.5% - 100%, however, showed more variability than the AE group range (95.8%-100%). As with the word-final results, Fisher's exact tests were conducted to determine if the group differences were significant. The difference between the PS and AE groups was significant (p = .01), but the difference between the HU and AE groups was not.

3.2.7. Word-initial fricative perception

Table 3.13 shows the perception results for word-initial fricatives across the three language groups. In the two-choice task, surprisingly it was the American English group that had the lowest perception accuracy scores. In the four-choice task, all groups showed lowered median accuracy scores.

Further analysis of the results of the two-choice task showed that the majority of errors made by the American English participants were in the voiced-voiceless (e.g. *Deb Simmer*) context (Median = 79.2%), followed by the voiceless-voiceless (e.g. *Depp Simmer*) context (Median = 91.7%). The other two contexts resulted in perfect accuracy by the American English participants. Both the Hungarian and Polish participants were highly accurate across all contexts. All contexts had median accuracy scores of over 98%, with the exception of the voiced-voiceless context (Median = 92.4% for both groups).

The fact that the voiced-voiceless context was the least accurate for all three language groups warranted further consideration. It is possible that the unnaturalness of the edited stimuli created perceptual confusion – recall that the stimuli were spliced *before* the release of the initial fricative. Amplitude ramps were included to minimize the sound of the preceding stop, but it is possible that the voiced stops were still audible in these tokens.

The four-choice context, however, would not contain any artifacts of splicing, so any errors in this context would need to be interpreted in another way. As in the two-choice task, the majority of errors were on the perception of word-initial /s/. The least accurate context was again the voiced-voiceless context, with all groups receiving low median scores (AE: 37.5%, HU: 58.3%, PS: 50.0%). The American English participants in addition produced low scores (Median = 66.7%) in the voiceless-voiceless context.

Since the goal of this study was to determine the effect of one's native phonological rules on perception, the preferred interpretation of these results would be one that takes into account how each group *produces* these sequences. This type of interpretation is explored in the discussion section. Since the American English results here are so striking, however, it is possible that there was a potential confound in the experimental design. It could be that the lower accuracy for *Simmer* is due to a name familiarity effect. It could be that the name *Zimmer* is much more frequent that *Simmer*, and therefore listeners were biased in favor of hearing /z/. Since *Zimmer/Simmer* was the only minimal word-initial /z/ - /s/ pair used in the perception study, this familiarity hypothesis cannot be tested.

3.3. Perception-Production Relation

Spearman rank order correlations were performed to determine how participants' productions related to their performance on the perception task. The overall word-final percent correct score on the four-choice task was used as the perception variable. For the production variable, each utterance was given a score based on its similarity to native productions. Those utterances which had word-final voicing values within the 25th – 100th percentiles of the native American English values were scored as correct. Those that were below the 25th percentile were scored as incorrect. For example, the score at the

25th percentile of AE voicing in *Buzz Potter* was 32%. So for non-native productions of *Buzz Potter*, those utterances that had 32% voicing or higher were scored as correct. Note that for voiced obstruents, the 100th percentile would be a high number (i.e. most voiced) while for voiceless obstruents, the 100th percentile would be a low number (i.e. least voiced). All the correct scores were totaled and then averaged together, resulting in one overall production score for each participant. With this method of scoring, native Hungarian speakers had a median production score of 54% correct, and native Polish speakers had a median production score of 45% correct.

Figure 3.4 shows a scatterplot diagram of the relation between overall perception scores to production scores. There was a slight positive correlation (Spearman's $\rho = .488$, $p < .01$, one-tailed). This suggests that those participants with the most native-like final voicing values were more likely to have higher perception scores, and vice versa. While it could be possible that the significant correlation is simply a reflection of the fact that the native Hungarians had higher perception and production scores overall, it should be noted that a native Polish speaker had the highest production score, and was tied for second with a native Hungarian.

Chapter 4. Discussion

4.1. Summary of results

In this section, the results of the production and perception tasks are summarized and discussed in terms of their support for the predictions discussed in the introduction. The perception results are discussed in terms of L1 phonetic versus phonological interference hypotheses presented in the introduction. Again, the predictions made by each of these hypotheses are being used to serve as extreme, idealized cases – with the expectation that the actual results will lie somewhere in between.

4.1.1. Word-final production

It was predicted, based on earlier research by Rubach (1984) and Altenberg and Vago (1983) that native Hungarian and Polish speakers would transfer their L1 regressive voicing assimilation rule into their English productions. That is, in sequences with a voicing mismatch (e.g. *Jez Keller*), we expected the non-native speakers to assimilate the voicing of the word-final obstruent to that of the following initial obstruent (*Jez Keller* → *Je[s] Keller*). Based on previous evidence from Cebrian (2000), which explored the transfer of Catalan voicing rules into L2 English productions, it was predicted that regressive *devoicing* would be more common than regressive *voicing*. The production results here show this to be the case. Both native Hungarian and native Polish listeners consistently devoiced word-final obstruents which preceded word-initial voiceless consonants. This was in contrast to native English speakers, who did not devoice stops in the same context, and only partially devoiced fricatives.

The tendency for native Hungarian and Polish speakers to voice word-final obstruents preceding voiced word-initial consonants was also significant, but not as strong as the tendency to regressively devoice. That is, while word-final voiceless obstruents were more voiced when in voiceless-voiced than voiceless-voiceless contexts, this tendency was more variable across participants, and less complete. Both non-native groups did, however, show more regressive voicing than native English speakers in the same contexts.

Consistent with predictions, word-final obstruents in sequences with matched voicing (e.g. *Jess Keller, Jez Geller*) were produced more successfully by the non-native groups than those with mismatched voicing. Word-final voiceless obstruents consistently remained voiceless in voiceless-voiceless sequences. Word-final voiced obstruents tended to remain voiced in voiced-voiced sequences, but they were devoiced more often by non-natives than by native English speakers.

Why is the tendency to devoice stronger than the tendency to voice in assimilation contexts? Cebrian (2000) observed this same pattern in Catalan late L2 learners of English, and attributed it to an interlanguage word-integrity effect, which restricts the application of rules to the level of the phonological word. That is, L1 rules are more likely to occur within a word boundary than across a word boundary. Under his account, all instances of word-final devoicing observed by Catalan L2 learners of English were instances of transfer of the L1 final obstruent devoicing rule, not the L1 regressive voicing assimilation rule. Final obstruent devoicing, being an extremely prevalent rule across the world's languages, creates an output which is unmarked. Transfer of final obstruent devoicing, then, is a reflection of universal tendencies. It could be, in accordance with Cebrian's findings, that the devoicing observed in the current study was due to transfer of final obstruent devoicing. The fact that both non-native groups sometimes devoiced fricatives even in voiced-voiced sequences provides support for this account. However, since only Polish, but not Hungarian, has a final devoicing rule in the L1, we would still need to explain why Hungarians had such a strong tendency to devoice. Altenberg and Vago (1983) observed this same pattern of final devoicing in English obstruents by Hungarian late L2 learners of English. Sentence-finally before a pause, both of their participants devoiced obstruents in words such as *colors, end,* and *increases*. Altenberg and Vago attributed the devoicing to the application of universal, unmarked rules. The findings in the current study, then, are consistent with their observations.

Regardless of which rule was transferred, the fact remains that both the Hungarian and Polish participants in this study had difficulty maintaining the English word-final voicing contrast in obstruent sequences crossing a word boundary. The issue then, is how to relate this difficulty in production to difficulties in word-final perception.

4.1.2. Word-initial production

It was predicted that native Hungarian and Polish speakers would produce word-initial stops with VOT values consistent with L1 VOT values: negative VOT for voiced stops, and short-lag VOT for voiceless stops. While both non-native groups produced voiced stops with negative VOT, they did not do so for every stop token in the sample. Productions of word-initial /g/ had fewer instances of negative VOT than productions of word-initial /b/. This pattern was consistent across all language groups, including the American English group. It appears that the non-native speakers in this sample were able to produce English short-lag voiced stops at least some of the time. Voiceless stops, as predicted, were produced with shorter VOT values by the Hungarian and Polish groups than the native American English group.

As for fricatives, word-initial /s/ was produced by all language groups with little or no voicing. Word-initial /z/ was, for the most part, produced with voicing throughout most of the frication interval. In the voiceless-voiced sequence *Buck Zelder*, however, /z/ was significantly more devoiced by native Hungarian and native English speakers than it was in the corresponding voiced-voiced context. The progressive devoicing observed here by native English speakers is not necessarily surprising, given that English has progressive devoicing as a morphological alternation in the plural (e.g. *cat[s]* versus *dog[z]*). I have not found any acoustic studies of Hungarian which show evidence of progressive devoicing, so it is unclear if the findings here reflect the acquisition of the English pattern or interference from Hungarian. The important issue for this study, however, is how this might affect word-initial fricative perception.

4.1.3. Word-final perception: two-choice versus four-choice task

In the perception task, participants identified word-final obstruents both before an inserted pause and before an obstruent. Results were predicted to pattern in one of two ways. If L1 interference operates strictly at a phonological level, we would expect word-final perception to be poor in pre-obstruent position in the four-choice task for both non-native groups. This context would be more difficult since L1 voicing assimilation rules should interfere with perception. Under this phonological account, Hungarian listeners should perform more accurately in the two-choice task, where the word-final obstruent precedes an inserted pause, since word-final voicing contrasts occur before a pause in

Hungarian. Polish listeners, on the other hand, should continue to do poorly due to transfer of their final obstruent devoicing rule.

If, however, L1 interference operates solely at a language-specific phonetic level, then perceptual difficulty should rest on prior L1 experience attending to word-final voicing cues. Hungarian listeners should perceive the target segments accurately regardless of whether they precede an inserted pause or an obstruent, since the exact same phonetic cues are available in both contexts. Polish listeners, on the other hand, should have difficulty with the target segments in both contexts due to their lack of L1 experience attending to word-final voicing cues.

The perception results did not confirm either one of the above scenarios. In the two-choice task, where the word-final obstruents preceded an inserted pause, the two non-native groups performed well above chance, but significantly more poorly than the native English speakers. In the four-choice task, where the word-final obstruents directly preceded an adjacent obstruent, identification accuracy by both non-native groups was significantly poorer than in the two-choice task. The accuracy rates of the Polish group were significantly lower than those of the Hungarian group only in the four-choice task.

Why were the non-native groups not significantly different in their perception of voicing contrasts in the two-choice task? Both scenarios above predicted that the native Hungarian listeners should be able to discriminate English word-final voicing contrasts, so their high accuracy levels were not surprising. The accuracy levels of the Polish listeners, however, were not predicted. Nor are they completely surprising, however. Earlier research demonstrated that speakers of languages without word-final voicing contrasts can perceive English word-final obstruents (Flege, 1989; Flege and Wang, 1989; Broersma, 2005) in isolated syllables in clear speech. In the present experiment, the speaker was instructed to speak naturally, but clearly. All word-final stops were produced with release bursts. The cues available in this clear speech mode may have made the voicing distinction more salient, even for the native Polish listeners. But even with the increased salience, neither non-native group was as accurate as the American English group. Native English speakers reported that the artificially inserted silence following the target name did cause the stimuli to sound less natural, and this may have affected perception. In addition, the presence of a carrier phrase, which creates a higher memory load, might have made the stimuli more challenging than a simple citation-style utterance.

The fact that both groups had reduced accuracy in the four-choice task was predicted by the phonological interference hypothesis, since both groups have L1 regressive voicing rules that were predicted to interfere with perception. It was not predicted by the phonetic interference hypothesis, since the same acoustic cues were available in both the two-choice and four-choice tasks. The fact that the Hungarian group was more accurate than the Polish group was more consistent with the phonetic interference hypothesis than the phonological interference hypothesis. But before we consider the relative merits of the two hypotheses, we need to consider the role that task complexity may have played in the results. The four-choice task is more complex than the two-choice task, since participants were required to identify both the final and initial obstruent simultaneously. The additional complexity, however, did not create difficulty for the American English participants. In fact, in two out of the three manner conditions, accuracy rates were better in the four-choice task for native listeners. Perhaps, then, task complexity was only a factor for non-native speakers. It could be that a challenging non-native contrast becomes even more challenging as task demands rise. This is indeed a possibility, but even if task complexity were a contributing factor, the higher accuracy of the native Hungarian group would still need to be explained.

The higher accuracy of the Hungarian group could, of course, be explained by the fact that Hungarian has a word-final voicing contrast before a pause. L1 experience attending to word-final cues to a voicing contrast in pre-pausal position may provide these listeners with an advantage in detecting word-final voicing contrasts in an assimilation context. This interpretation would give more support to the phonetic interference hypothesis. However, the fact that the word-final accuracy levels of native Hungarians in this sample decreased in the four-choice task suggests that the pre-obstruent position in itself created difficulty. So it seems plausible that phonological interference played a role in the observed accuracy levels.

Based on the complex and unexpected pattern of results, it seems possible that all three factors – phonological interference, phonetic interference, and task complexity – affected listeners. It could be that only when you push the limits of perceptual abilities through task complexity that you see the relative effects of L1 experience. That both groups did well in the two-choice task suggests that both could utilize the available phonetic cues to the voicing distinction in word-final utterances produced in the assimilation context. That both groups did more poorly in the four-choice task implies that the presence of a following obstruent interfered with perception, likely due to the effect of transfer of L1 regressive voicing assimilation rules in both Hungarian and Polish. And finally, that Hungarians did better than Polish implies that their experience with cues to L1 voicing contrasts before a pause aided them when L2 voicing contrasts were in pre-obstruent position. It appears, then, that L2 learners are affected by prior experience with rules as well as individual phonetic cues, and it is the manipulation of task variables that reveal where each of these levels comes into play. Further research is needed to tease these factors apart.

4.1.4. Word-final perception across voicing contexts

The relative support for the phonological versus phonetic hypotheses can be further tested by examining word-final perception across the different voicing contexts. Under an L1 phonological interference hypothesis, all four voicing contexts (voiced-voiced, voiced-voiceless, voiceless-voiced, voiceless-voiceless) should be equally difficult. If one's L1 assimilation rule creates the expectation that the second member of an obstruent sequence dictates the voicing of the entire sequence, then listeners may have the tendency to interpret the surface form as non-underlying. Under an L1 phonetic interference hypothesis, the manner in which word-final obstruents are perceived in the two-choice task should be the same as they are in the four-choice task. This has already been shown to be incorrect.

The results for perception in the different voicing contexts did not exhibit any one clear pattern. Polish participants performed significantly more poorly than native AE participants in *all* contexts overall, so there was no one context that stood out as more successfully perceived than the others. While that result might appear to support the phonological interference hypothesis, the accuracy rates across the individual manner conditions tell a different story. Polish participants did have accuracy rates as high as the native English listeners in a few contexts -- voiced-voiced and voiceless-voiceless in fricative-stop sequences, and voiced-voiced in stop-fricative sequences. The Hungarian results also failed to exhibit any consistent pattern. Overall and across the individual manner conditions, there was no one context where they were consistently less accurate than the native American English participants. The only clear result is that native Hungarians were consistently as accurate as the native English listeners in the voiced-voiced context. Since there are some contexts that were more accurate than others, this is counter-evidence for the pure form of the phonological interference hypothesis.

Among these somewhat inconsistent results, the clearest finding is that the voiced-voiced context was the most accurate for both non-native groups. Hungarians were most accurate in this context overall as well as across the different manner conditions, and Polish listeners were most accurate in this context in two manner conditions. Why might this be? Looking at the acoustic measurements of the stimuli, word-final voiced segments in this context were always fully voiced. It could be, then, that word-final voiced obstruents were easier to detect in the voiced-voiced context because they were more fully voiced than they were in the comparable voiced-voiceless context. Word-final voiced fricatives in the voiced-voiceless context were indeed partially devoiced. But word-final stops in the same context were all fully voiced except for two tokens of *Mag* (which were at least 70% voiced). If the strength of phonetic voicing cues were the sole reason for the accurate detection of word-final voicing, you would expect that the stops in the voiced-voiceless context would be perceived more accurately than the fricatives, since it was the fricatives that were partially devoiced. This, however, is not reflected in the perception scores from the non-native listeners. If anything, the word-final stops were less accurately perceived than the fricatives in this context. It seems, then, that the success with word-final voiced obstruents in the voiced-voiced context can be attributed to the presence of the following obstruent. One interpretation as to why word-final voicing is more perceptible before a voiced obstruent could be that the voicing cues across a matched voiced-voiced sequence reinforce each other. Another interpretation could be that matched voicing sequences, being less marked than mismatched voicing sequence, are easier to perceive. Neither of these interpretations, however, explain why a voiced-voiced sequence should be *better* than a voiceless-voiceless sequence. If voicing cues across matched voicing sequences perceptually reinforce each other, then voiceless-voiceless sequences should be perceived just as accurately than voiced-voiced sequences. Similarly, if matched voicing sequences are easier due to the fact that they are unmarked, then voiceless-voiceless sequences should be easier, as well.

The phonological interference hypothesis may provide some insight into this problem. We saw from the production data that the non-native speakers had a strong tendency to devoice voiced obstruents, especially when they preceded voiceless obstruents. Perception of voicelessness, then, is not assurance that a segment is underlyingly voiceless. Listeners do not know if a segment is voiceless because it has assimilated, or because it is underlyingly voiceless. Interestingly, the one context that American English listeners received the lowest median accuracy scores was in voiceless-voiceless sequences in the fricative-stop condition (Median = 83.3%). As noted earlier, voiced fricatives were partially devoiced in both the stimuli and the AE participants in the production task in the voiced-voiceless context. It is possible, then, that AE listeners, knowing that voiced fricatives can be devoiced before a voiceless consonant, have trouble interpreting voicelessness in fricatives in this position. Similarly, the non-natives, who tend to devoice segments in all contexts, are less willing to say a word-final segment is voiceless than they are to say it is voiced when they hear positive evidence of voicing.

4.1.5. Word-initial stop perception

Word-initial stop perception was highly accurate for all language groups in the two-choice task. Accuracy rates were slightly lower in the four-choice task for the non-native speakers, and the difference between the native English speakers and both the Hungarian and Polish groups was significant. It could be, again, that the increased task complexity was responsible for the decreased accuracy levels of non-native listeners to decrease. However, if task complexity alone were the sole cause of the lower accuracy levels, one would not expect any one context to be more difficult than others. As it was, not all contexts were perceived equally well.

It was hypothesized that voiced stops might be difficult for the non-native groups when preceded by a voiceless obstruent. In this context, the short-lag VOT of the voiced stops could have been confused for a voiceless stop, since voiceless stops in Hungarian and Polish have short-lag VOT values. In addition, the mismatch in voicing between the final and initial consonant could cause confusion for the non-natives – the voicelessness of the word-final consonant might lead to an expectation that the following obstruent was voiceless, as well. This hypothesis was partially supported: native Polish listeners did perform significantly more poorly than native English listeners on voiced stops in voiceless-voiced contexts. Native Hungarian listeners did not perform significantly more poorly than the native English listeners in this context, but it should be noted that three of the 11 listeners had accuracy levels at 66.7% for stops in this context, which was far lower than accuracy rates for any other initial stop in any context for the Hungarian group. So it appears this context was problematic for some, but not all Hungarian listeners.

The reason for the difficulty cannot be simply that the stops were miscategorized due to their phonetic similarity to Hungarian and Polish voiceless stops. If short-lag VOT was the sole problem, we would expect to see the same accuracy rates in the two-choice task, and this was not the case. It appears then, that perception of the word-initial voiced stop is affected by the preceding voiceless obstruent. It could be that the first voiceless segment in the sequence causes listeners to anticipate voicelessness in the following segment, in accordance with L1 assimilation rules. This anticipatory effect has been observed in previous research (Gaskell and Marslen-Wilson, 1998; Gow, 2001, 2003; Otake, et al., 1996). This result, then, gives support to the hypothesis that L1 phonological interference affects perception of L2 contrasts.

It should be noted that in terms of mean accuracy levels, the least accurate voicing context for initial consonants was, as describe above, the voiceless-voiced context (HU: 87.1%, PS: 87.5%). The second least accurate was the voiced-voiceless context (HU: 95.8%, PS: 94.4%). The fact that the two mismatched voicing contexts were the least accurate gives further support to the L1 phonological interference hypothesis. However, since accuracy rates were so high, it is difficult to say if this is a significant finding. Reaction time studies, which more clearly show processing effects when perceptual performance is at ceiling, might be useful if this issue were to be explored further.

4.1.6. Word-initial fricative perception

As discussed in Chapter 3, the results for word-initial fricative perception were surprising. While initial /z/ was perceived highly accurately by all language groups, word-initial /s/ perception was perceived significantly less accurately. Native English listeners had the most difficulty with /s/, obtaining the lowest median accuracy scores of the three language groups. Since word-initial /s/ was voiceless in the stimuli, it is not clear why this segment was so difficult to perceive. In Chapter 3 the possibility of name familiarity was discussed – it could be that the frequency of the name *Zimmer* biased participants into mislabeling /s/ as /z/. Since *Zimmer/Simmer* was the only pair used with initial fricatives, this possibility cannot be disproved with this set of data.

Participants' word-initial fricative productions could potentially provide some insight into their perceptual performance. Native English speakers produced word-initial /z/ in one voiceless-voiced context (*Buck Zelder*) with significantly less voicing than in the comparable voiced-voiced context. Hungarian speakers also showed devoicing of word-initial voiced fricatives, with more devoicing overall than the other two language groups. Interestingly, it is the English and Hungarian groups that have the lowest scores on word-initial fricative perception (AE: 77.5%, HU: 77.1%, PS: 87.5%). It could be that the pattern of devoicing observed in the production data might be related to the errors in the perception data. The American English group misperceived word-initial /s/ even when it followed a voiceless obstruent (Median = 66.7%). This is a potential parallel to the situation with word-final voiceless-voiceless sequences discussed in section 4.1.4. It could be that since the listeners know that progressive devoicing does occur in these sequences, they may not interpret fricatives in a voiceless-voiceless context as underlyingly voiceless. Further research would need to confirm whether the pattern here reflects something more interesting with respect to native language phonotactics, or whether it is simply a lexical frequency effect.

4.2. Perception-Production relation

The correlation between overall non-native perception and production scores was shown to be significant, accounting for 24% of the variance. Further research with a larger sample of participants is needed to show if this was simply a reflection of the fact that the native Hungarian group had higher production and perception scores than the native Polish group. The pattern of perception results discussed above, however, appear to reflect what the participants were doing in production. The finding that word-final perception was more accurate in the voiced-voiced than the other three contexts could be explained with reference to the fact that regressive devoicing was a more common production error than regressive voicing. Similarly, the surprising finding that /s/ was poorly perceived by American English and Hungarian speakers had a potential explanation in the fact that progressive devoicing was observed in American English and Hungarian productions. Again, while further research would confirm the strength of these observations, at the very least the perception results show that these listeners do have difficulty with a contrast that is difficult in production.

4.3. Implications for theoretical models

The production results here are consistent with predictions of the Contrastive Analysis Hypothesis (CAH) and the Markedness Differential Hypothesis (MDH). The CAH predicts that the presence of a rule in Hungarian and Polish that is absent in English will cause difficulty for L2 learning. It does not predict the direction of difficulty (i.e. whether the presence of the rule should be problematic for Hungarian and Polish learners of English, or vice versa). The MDH fills in the directionality gap – it predicts that the regressive voicing assimilation rule in Hungarian and Polish, being an unmarked process, should be difficult to suppress when learning English. Both the CAH and MDH, then, correctly predict this area of difficulty for Hungarian and Polish learners of English.

Similarly, with respect to perception, the MDH correctly predicts that English word-final voicing contrasts in pre-obstruent position will be difficult for Hungarian and Polish learners of English. Since the English contrast does not exist in Hungarian or Polish in pre-obstruent position, it would be considered marked and therefore more difficult to acquire.

The SLM and PAM also make predictions about performance in L2 perception. The SLM states that L2 learners should be able to create categories for 'new' sounds, but similar sounds will continue to be assimilated to L1 categories. In the case of the English word-final voicing contrasts examined in the current study, the categories are not new for Hungarian or Polish listeners, since both Hungarian and Polish do have word-final obstruents. It is likely, then, that Hungarian and Polish L2 learners of English will assimilate the English phones to L1 categories. Since the L1 speech sounds are non-contrastive in the L1, they may continue to be difficult to produce and perceive in the L2. Under this interpretation, the SLM would accurately predict the basic finding that the participants in this study had difficulty with the English word-final voicing contrast.

To interpret the perception results according to the predictions of PAM, it is important first to determine how PAM would define categories for word-initial versus word-final obstruents. If we assume that the same category is established for word-final obstruents as word-initial obstruents, then PAM would predict that perception of the

English word-final voicing contrast should be good, since both Hungarian and Polish have a word-initial voicing contrast. Broersma (2005) interpreted PAM in this way, confirming her finding that Dutch listeners were successful at the perception of the English word-final voicing contrast. The results observed in the current study would not support PAM under this interpretation, since word-initial perception was much more successful than word-final perception.

If we assume, however, that categories are established at the level of the positional allophone, then perception by L2 learners would depend on how well word-final obstruents were assimilated to native categories. In a context where a given contrast is neutralized in the L1, we would expect that the members of the contrast in the L2 would be assimilated to the same L1 category. Perception could potentially range from poor (single category assimilation pattern) to good (category goodness pattern), depending on the salience of the phonetic cues to the distinction. The assimilation of English voicing contrasts to native Hungarian and Polish categories was not directly assessed in the current study, but the fact that the participants found the contrast difficult seems to imply that these contrasts conform to a single category assimilation pattern.

What is more difficult to address using PAM or the SLM, however, is the different results across the two different tasks (two- versus four-choice) and voicing contexts. It would be counterproductive to have to assess perceptual assimilation for all contrasts in all contexts to predict results. A recent model of L1 and L2 speech, Automatic Selective Perception (ASP, Strange and Shafer, 2008) provides a framework from which to address the potentially puzzling pattern of results across different stimulus and task variables. According to ASP, perception of native contrasts is predicted to be robust even in complex tasks and non-optimal listening conditions. Perception of non-native contrasts is predicted to vary based on the nature of the task. Simple tasks which require few cognitive resources are predicted to tap into a universal phonetic mode of processing. Perception of non-native contrasts in these tasks is predicted to be good. More complex tasks are predicted to tap into a language-specific, phonological mode of processing. Perception of non-native contrasts in these tasks is predicted to be less accurate. These predictions are confirmed by the current study. The native English participants had high accuracy rates in both the two- and four-choice tasks. Any increased cognitive load imposed by the four-choice task did not hinder L1 perception. The non-native participants, however, had lower accuracy rates in the four-choice task than in the two-choice task. It seems likely that the two-choice task, which contained simpler stimuli and response options, was able to tap into the universal phonetic mode of processing. That the two non-native groups did equally well on this task despite different L1 experience with word-final voicing contrasts supports this position further. Had citation style utterances been used, like those used in Flege (1989), Flege and Wang, and Broersma (2005), perception likely would have been even more accurate. The four-choice task, however, not only required a higher cognitive load (due to the fact that both the initial and final obstruent needed to be identified), but also required the suppression of native phonological processes. As discussed earlier, further research is needed to tease apart the relative contributions of each of these factors (task complexity versus phonological interference). In either case, however, ASP correctly predicts the pattern of results observed here.

4.4. Conclusions and directions for future research

This study was designed to explore whether L1 phonological processes would interfere with the perception and production of voicing in a second language. The results here provided some evidence that the L1 phonological rules of final devoicing and regressive voicing assimilation transferred to the L2 productions of Hungarian and Polish late learners of English. L2 perception also appeared to be affected by native assimilation rules, as some of the results from the perception task could not be explained solely by reference to L1 context-independent phonetic interference. Hungarian and Polish listeners were less accurate at identifying word-final voicing contrasts in pre-obstruent position than they were at identifying those same exact contrasts in pre-pausal position. They were also slightly less accurate than native English listeners at identifying word-initial voiced stops in obstruent sequences with mismatched voicing, a finding which was not seen for the same sequences in the post-pausal condition.

These results are consistent with the notion that marked structures cause difficulty in the acquisition of a second language. Word-final voicing contrasts, which are typologically marked, were difficult for native Hungarian and Polish speakers to produce and perceive. Future research should explore the extent to which universal tendencies might further explain some of the more specific findings observed here, such the direction of difficulty observed across voicing contrasts, and the lower accuracy rates of American English participants on voiceless fricatives in certain contexts.

Further research is needed to replicate the findings here. The set of stimuli in any one study is necessarily restricted – replication with different lexical items would ensure that any results found here were not due to lexical familiarity effects. In the four-choice task, word-final and initial perception were assessed simultaneously, since it was not clear how each would be perceived in these contexts. Now that we know that word-final perception is significantly poorer than word-initial perception in obstruent-obstruent sequences, it would be useful to perform the same task again, this time assessing word-final perception in pre-obstruent position in a two-choice task. If participants still have reduced accuracy in pre-obstruent position, this will further support the conclusion that L1 phonological rules interfere with L2 perception.

5. Tables

Table 1.1

Summary of voicing in English, Hungarian, and Polish

	Word-initial voicing	Maintains word-final voicing contrast?	Maintains voicing contrast before obstruents?
English	Short-lag vs. long-lag VOT	Yes	Yes
Hungarian	Negative vs. short-lag VOT	Yes	No
Polish	Negative vs. short-lag VOT	No	No

Table 2.1

Perception stimuli, broken down by sequence type, word-initial sequence, and word-final sequence

Sequence Type	Word-initial sequence	Word-final sequence	First name	Last name
Stop-Stop	V_1K/V_1G (vel)	P/B (lab)	Mack/Mag	Pillman/Billman
Stop-Stop	V_2P/V_2B (lab)	K/G (vel)	Depp/Deb	Keller/Geller
Stop-Fric	V_1K/V_1G (vel)	S/Z (alv)	Mack/Mag	Simmer/Zimmer
Stop-Fric	V_2P/V_2B (lab)	S/Z (alv)	Depp/Deb	Simmer/Zimmer
Fric-Stop	V_1S/V_1Z (alv)	P/B (lab)	Kass/Kaz	Pillman/Billman
Fric-Stop	V_2S/V_2Z (alv)	K/G (vel)	Jess/Jez	Keller/Geller

Table 2.2

Summary of acoustic measurement data for perception stimuli – word final consonants

	Vowel Duration (ms)		Fricative Duration (ms)		% Voicing	
Consonant	Median	Range	Median	Range	Median	Range
/z/	139	118-171	47	50-75	78.3	23.1-100
/s/	126	113-165	78	61-96	8.1	0-16.2
/b/	101	89-114	-	-	100	100-100
/p/	85	76-92	-	-	27.6	0-67.8
/g/	130	113-153	-	-	100	70.5-100
/k/	122	108-140	-	-	0	0-20.9

Table 2.3

Summary of acoustic measurement data for perception stimuli – word-initial stops

	Lead VOT (ms)		Lag VOT (ms)	
Language	Median	Range	Median	Range
/b/*	-67	-116 - -33	8	8-10
/p/	-	-	49	32-73
/g/**	-48	-72 - -38	12	8-17
/k/	-	-	60	40-94

*18 tokens with lead VOT, 6 with lag
**13 tokens with lead VOT, 11 with lag

Table 2.4

Summary of acoustic measurement data for perception stimuli – word-initial fricatives

Language	Fricative Duration (ms)		% Fricative Voicing	
	Median	Range	Median	Range
/z/	85	54-137	100	12.5-100
/s/	120	103-150	0	0-15.0

Table 3.1

Word-final percent voicing across languages – production data.

		AE Percent Voicing				HU Percent Voicing				PS Percent Voicing			
		C2 Voiced		C2 Voiceless		C2 Voiced		C2 Voiceless		C2 Voiced		C2 Voiceless	
Obs.	Context	Med%	I-QRange	Med%	I-QRange	Med%	I-QRange	Med%	I-QRange	Med%	I-QRange	Med%	I-QRange
/z/	Buzz+Stop	**100**	100-100	**46**	32-55	**53**	19-100	**6**	0-9	**68**	23-100	**10**	6-17
	Liz+Stop	**56**	41-100	**31**	23-37	**37**	19-100	**8**	6-10	**32**	25-90	**14**	9-19
/s/	Gus+Stop	**9**	7-11	**7**	3-10	**13**	5-28	**2**	0-7	**10**	5-35	**6**	0-8
	Siss+Stop	**13**	10-20	**10**	7-12	**17**	9-31	**7**	6-10	**28**	13-49	**11**	8-14
/g/	Jug+Stop	**100**	81-100	**100**	55-100	**100**	88-100	**0**	0-0	**100**	28-100	**6**	0-20
	Jug+Fric	**100**	100-100	**100**	100-100	**100**	100-100	**0**	0-22	**100**	61-100	**22**	2-50
/k/	Buck+Stop	**0**	0-7	**0**	0-0	**55**	0-100	**0**	0-0	**35**	0-100	**0**	0-9
	Buck+Fric	**6**	0-38	**0**	0-15	**48**	0-100	**0**	0-0	**35**	0-100	**0**	0-19
/b/	Gib+Stop	**100**	100-100	**100**	100-100	**100**	56-100	**36**	20-51	**100**	58-100	**41**	29-54
	Gib+Fric	**100**	100-100	**100**	89-100	**100**	94-100	**45**	21-66	**100**	57-100	**50**	30-71
/p/	Kip+Stop	**35**	28-45	**29**	22-39	**47**	35-83	**27**	16-37	**48**	31-85	**31**	25-46
	Kip+Fric	**61**	22-100	**31**	15-41	**77**	39-100	**30**	17-40	**62**	37-100	**37**	25-50

Table 3.2

Median word-final fricative duration across language – production data

		AE Final Fricative Duration				HU Final Fricative Duration				PS Final Fricative Duration			
		C2 Voiced		C2 Voiceless		C2 Voiced		C2 Voiceless		C2 Voiced		C2 Voiceless	
Obs.	Context	Med (ms)	I-QRange	Med (ms)	I-QRange	Med (ms)	I-QRange	Med (ms)	I-QRange	Med (ms)	I-QRange	Med (ms)	I-QRange
/z/	Buzz+Stop	**61**	53-69	59	51-81	**69**	59-92	79	66-97	**75**	58-102	94	82-112
	Liz+Stop	**78**	69-91	81	75-95	**85**	74-94	110	101-116	**100**	86-112	125	103-144
/s/	Gus+Stop	**81**	73-93	76	71-95	**85**	69-101	93	81-109	**98**	70-126	103	78-127
	Siss+Stop	**97**	90-114	106	101-117	**98**	86-106	127	106-141	**95**	77-120	135	118-154

Table 3.3

Median word-initial percent voicing across languages – production data

		AE % Initial Fricative Voicing				HU % Initial Fricative Voicing				PS % Initial Fricative Voicing			
		C1 Voiced		C1 Voiceless		C1 Voiced		C1 Voiceless		C1 Voiced		C1 Voiceless	
Obs.	Context	Med%	I-QRange	Med%	I-QRange	Med%	I-QRange	Med%	I-QRange	Med%	I-QRange	Med%	I-QRange
/z/	VelStop+ Zelder	**100**	100-100	**56**	0-100	**79**	13-100	**32**	0-81	**100**	34-100	**85**	0-100
	LabStop+ Zelder	**100**	100-100	**100**	48-100	**41**	7-100	**25**	0-39	**100**	28-100	**100**	13-100
/s/	VelStop+ Simpson	**0**	0-0	**0**	0-0	**0**	0-0	**0**	0-0	**0**	0-0	**0**	0-0
	LabStop+ Simpson	**0**	0-10	**0**	0-0	**0**	0-0	**0**	0-0	**0**	0-0	**0**	0-0

Table 3.4

Median word-initial fricative duration across languages – production data

		AE Initial Fricative Duration				HU Initial Fricative Duration				PS Initial Fricative Duration			
		C1 Voiced		C1 Voiceless		C1 Voiced		C1 Voiceless		C1 Voiced		C1 Voiceless	
Obs.	Context	Med (ms)	I-QRange	Med (ms)	I-QRange	Med (ms)	I-QRange	Med (ms)	I-QRange	Med (ms)	I-QRange	Med (ms)	I-QRange
/z/	VelStop+ Zelder	**94**	79-101	**107**	90-122	**103**	92-115	**113**	104-130	**114**	93-129	**113**	93-134
	LabStop+ Zelder	**98**	82-114	**108**	96-119	**113**	92-125	**115**	93-137	**120**	100-132	**131**	103-141
/s/	VelStop+ Simpson	**134**	113-150	**133**	110-146	**134**	123-161	**137**	115-150	**146**	133-163	**147**	136-178
	LabStop+ Simpson	**134**	116-155	**135**	110-147	**142**	122-164	**135**	127-155	**143**	135-166	**149**	132-168

Table 3.5

Median vowel durations across languages

Consonant	AE Vowel Duration (ms)		HU Vowel Duration (ms)		PS Vowel Duration (ms)	
	Median	I-Q Range	Median	I-Q Range	Median	I-Q Range
/z/	**123**	108–139	**121**	103–141	**111**	94–143
/s/	**91**	68–107	**108**	94–137	**93**	73–123
/b/	**92**	75–104	**91**	76–105	**83**	62–104
/p/	**61**	49–71	**67**	60–77	**53**	38–70
/g/	**120**	112–137	**131**	118–142	**124**	108–153
/k/	**97**	87–106	**124**	107–133	**113**	98–132

Table 3.6

Median VOT durations across languages

	AE VOT (ms)			HU VOT (ms)			PS VOT (ms)		
	N	Median	I-Q Range	N	Median	I-Q Range	N	Median	I-Q Range
/b/-Lead	20	**-91**	-102 – -66	48	**-96**	-120 – -73	74	**-86**	-98 – -80
-Lag	44	**8**	7–10	40	**8**	7–12	22	**11**	7–12
/p/-Lag	64	**52**	29–66	88	**18**	11–31	96	**22**	14–40
/g/-Lead	12	**-53**	-58 – -47	23	**-53**	-71 – -39	51	**65**	-73 – -52
-Lag	52	**19**	15–24	65	**23**	19–26	45	**23**	17–27
/k/-Lag	64	**64**	56–75	88	**45**	32–53	96	**49**	37–65

Table 3.7

Median word-final percent accuracy before an inserted pause – two-choice task

	American English		Hungarian		Polish	
	Median%	Range	Median%	Range	Median%	Range
Fricative-Stop	**96.9**	85.4-100	**91.7**	68.8-97.9	**87.5**	72.9-100
Stop-Stop	**95.8**	85.4-97.9	**91.7**	66.7-100	**89.6**	62.5-97.9
Stop-Fricative	**93.8**	89.6-100	**83.3**	72.9-100	**86.5**	70.8-100
Overall	**94.4**	91.7-97.9	**87.5**	71.5-96.5	**87.2**	72.2-95.1

Table 3.8

Median word-final percent accuracy before an obstruent –four-choice task

	American English		Hungarian		Polish	
	Median%	Range	Median%	Range	Median%	Range
Fricative-Stop	**93.8**	79.2-100	**89.6**	60.4-100	**72.9**	62.5-89.6
Stop-Stop	**97.9**	93.8-100	**83.3**	64.6-97.9	**70.8**	56.2-95.8
Stop-Fricative	**97.9**	87.5-100	**83.3**	60.4-91.7	**68.7**	60.4-89.6
Overall	**95.8**	90.3-98.6	**87.5**	66.7-93.8	**71.2**	61.8-91.7

Table 3.9

Overall median word-final percent accuracy across voicing contexts – four-choice task

	American English		Hungarian		Polish	
	Median%	Range	Median%	Range	Median%	Range
Vd Vd	**100**	86.1-100	**97.2**	63.9-100	**77.8***	61.1-100
Vd Vls	**94.4**	80.6-100	**86.1***	27.8-97.2	**55.6***	38.9-86.1
Vls Vd	**95.8**	86.1-100	**86.1**	50.0-100	**69.4***	50.0-86.1
Vls Vls	**94.4**	88.9-100	**88.9***	58.3-97.2	**84.7***	55.6-97.2

*Significantly lower ($p < .05$, one-tailed) than the AE score for the same context in median test

Table 3.10

Word-final percent accuracy broken down by name

		Before an inserted pause (two-choice task)		Before an obstrent (four-choice task)	
		Median %	I-Q Range	Median %	I-Q Range
Vd	Deb	**91.7**	79.2-95.8	**83.3**	75.0-91.7
Vls	Depp	**87.5**	58.3-100	**83.3**	54.2-95.8
Vd	Mag	**95.8**	95.8-100	**91.7***	62.5-100
Vls	Mack	**100**	95.8-100	**95.8***	91.7-100
Vd	Jez	**100***	91.7-100	**91.7***	83.3-100
Vls	Jess	**91.7***	75.0-100	**83.3***	66.7-100
Vd	Kaz	**91.7**	83.3-100	**91.7**	75.0-100
Vls	Kass	**91.7**	75.0-100	**91.7**	83.3-100

*significantly different ($p < .05$, two-tailed) from the other member of the minimal pair in that context in a Wilcoxon signed-ranks test

Table 3.11

Median word-initial stop percent accuracy after an inserted pause – two-choice task

	American English		Hungarian		Polish	
	Median%	Range	Median%	Range	Median%	Range
Fricative-Stop	**100**	100-100	**100**	95.8-100	**100**	100-100
Stop-Stop	**100**	100-100	**100**	93.8-100	**100**	97.9-100
Overall	**100**	100-100	**100**	94.8-100	**100**	99.0-100

Table 3.12

Median word-initial stop percent accuracy after an inserted pause – four-choice task

	American English		Hungarian		Polish	
	Median%	Range	Median%	Range	Median%	Range
Fricative-Stop	**99.0**	97.9-100	**97.9**	89.6-100	**95.8**	89.6-100
Stop-Stop	**100**	97.9-100	**95.8**	83.3-100	**94.8**	83.3-100
Overall	**99.0**	97.9-100	**93.8**	88.5-100	**95.8**	86.5-100

Table 3.13

Median word-initial fricative percent accuracy

	American English		Hungarian		Polish	
	Median%	Range	Median%	Range	Median%	Range
Two-Choice	**92.7**	77.1-97.9	**97.9**	93.8-100	**97.9**	93.8-97.9
Four-Choice	**75.0**	54.2-95.8	**77.1**	54.2-95.8	**87.5**	70.8-91.7

6. Figures

Figure 2.1

The following figure shows both the waveform and spectrogram of [ʤɛz#g] extracted from the sentence 'I met Jez Geller today'. Regions are as follows: 1) Vowel duration, 2) Fricative Duration 3) Closure Duration. Voicing is present throughout the entire closure, which is measured as negative (lead) VOT.

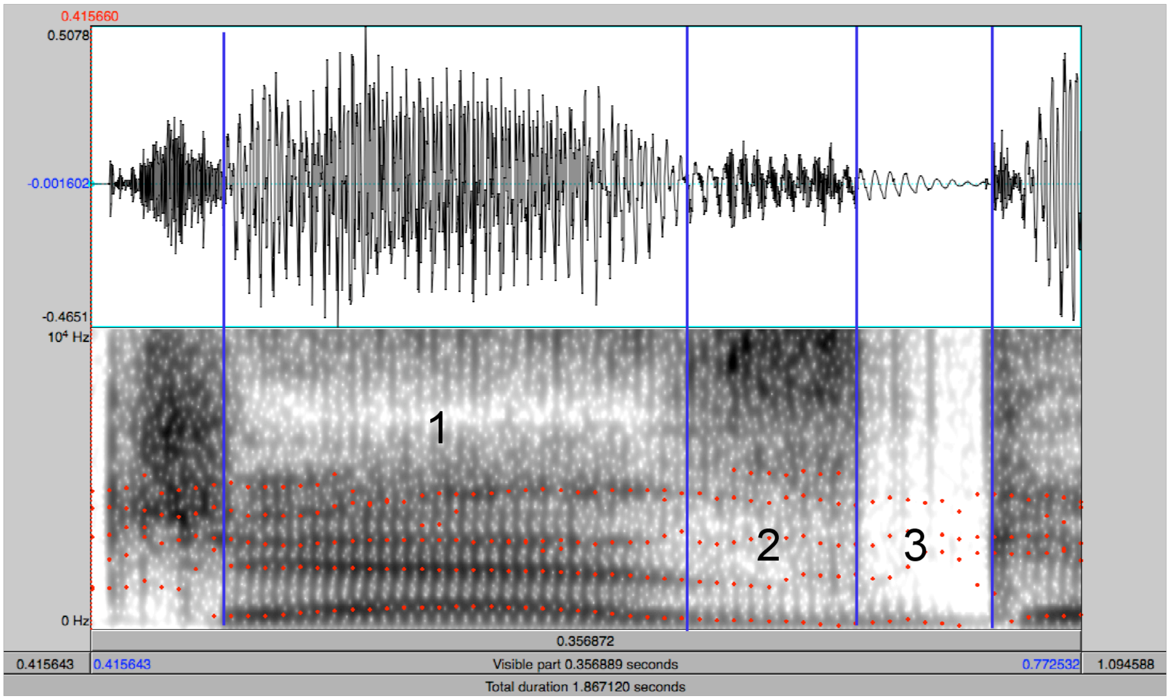


Figure 2.2

The following figure shows both the waveform and spectrogram of [dɛb#k^h] extracted from the sentence 'I met Deb Keller today'. Regions are as follows: 1) Vowel duration, 2) Closure 1 Duration 3) Release burst + low amplitude noise duration 4) Closure 2 duration 5) Positive VOT. Voicing is present throughout the entire duration of Closure 1.

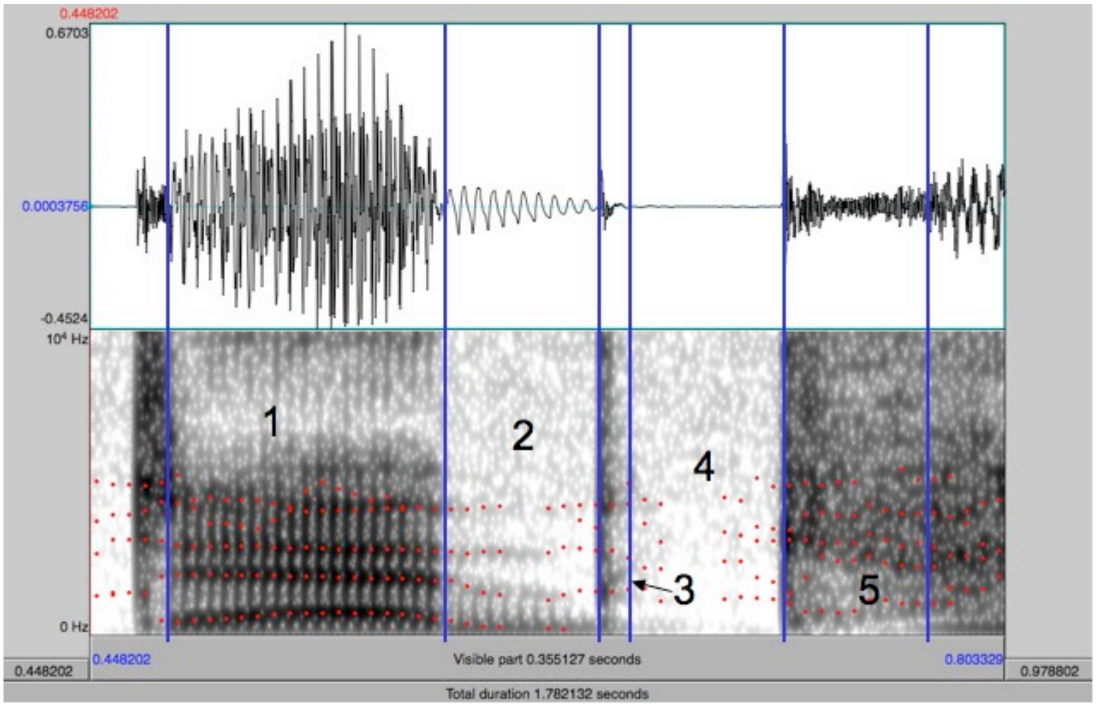


Figure 2.3

The following figure shows both the waveform and spectrogram of [mæg#s] extracted from the sentence 'I met Mag Simmer today'. Regions are as follows: 1) Vowel duration, 2) Closure Duration, 3) Release Burst + low amplitude noise duration, 4) Fricative Duration. Voicing is present throughout the entire closure duration.

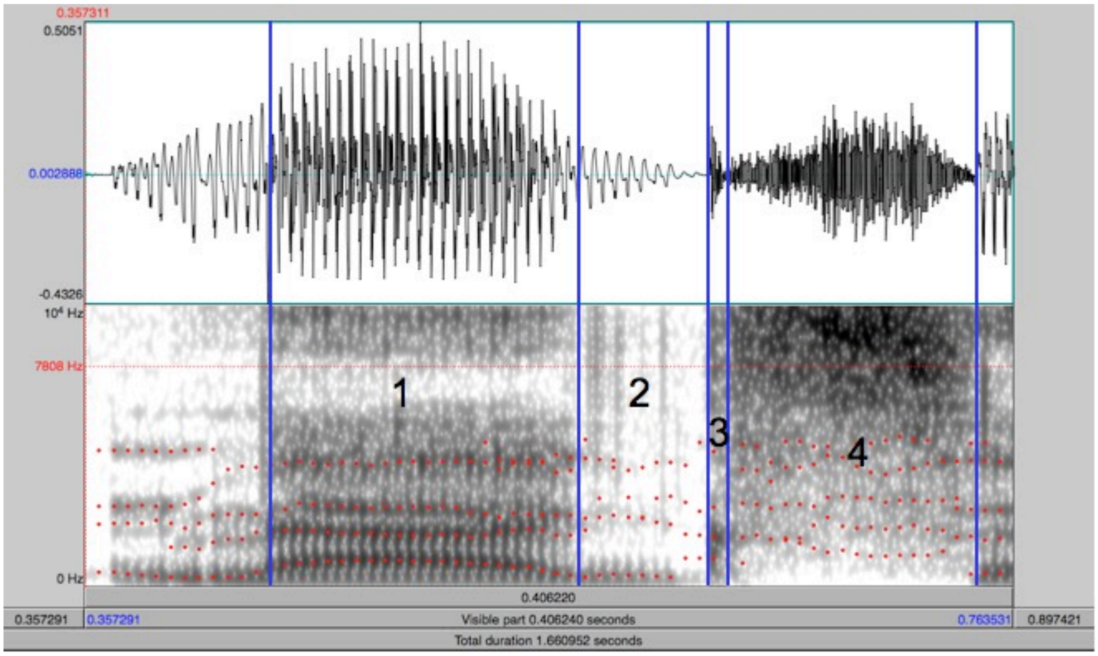


Figure 3.1

Spectrograms illustrating portions of the name Jug Potter, produced by two different native American English speakers. The /g/ in the top panel is released, resulting in two closures. The /g/ in the bottom panel is unreleased, resulting in one single closure. This difference in realization of the stop release makes it difficult to compare word-final stop voicing.

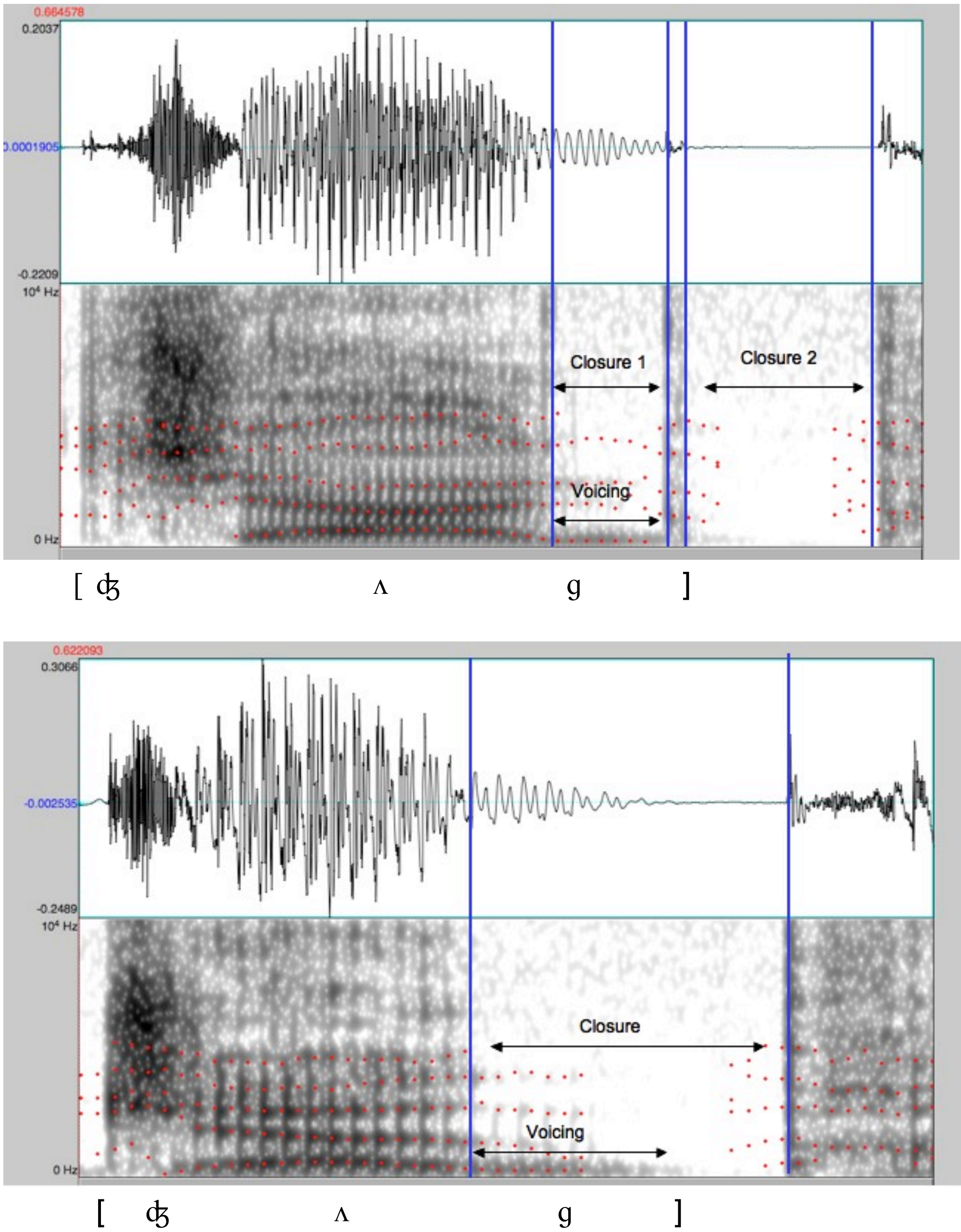


Figure 3.2

Percent word-final correct overall before an inserted pause – two-choice task

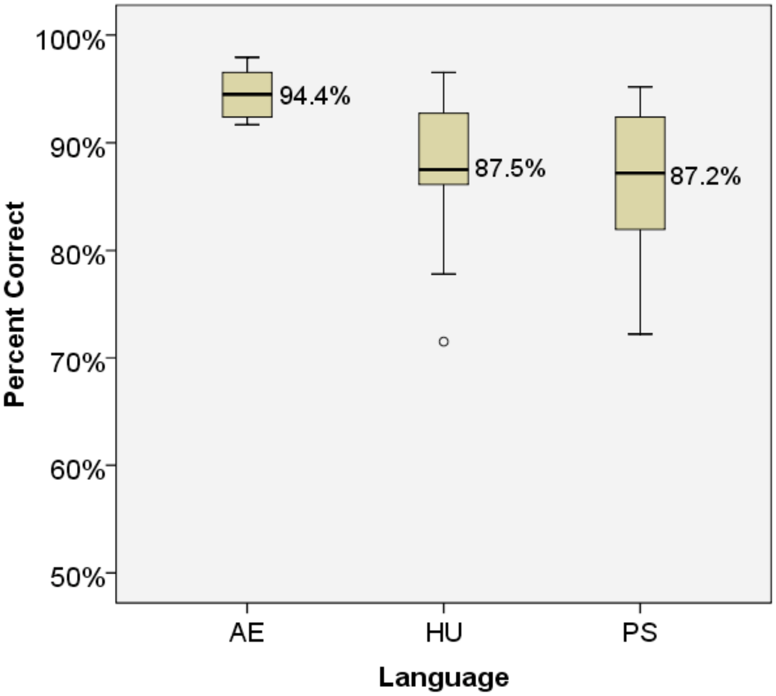


Figure 3.3

Percent word-final correct overall before an obstruent – four-choice task

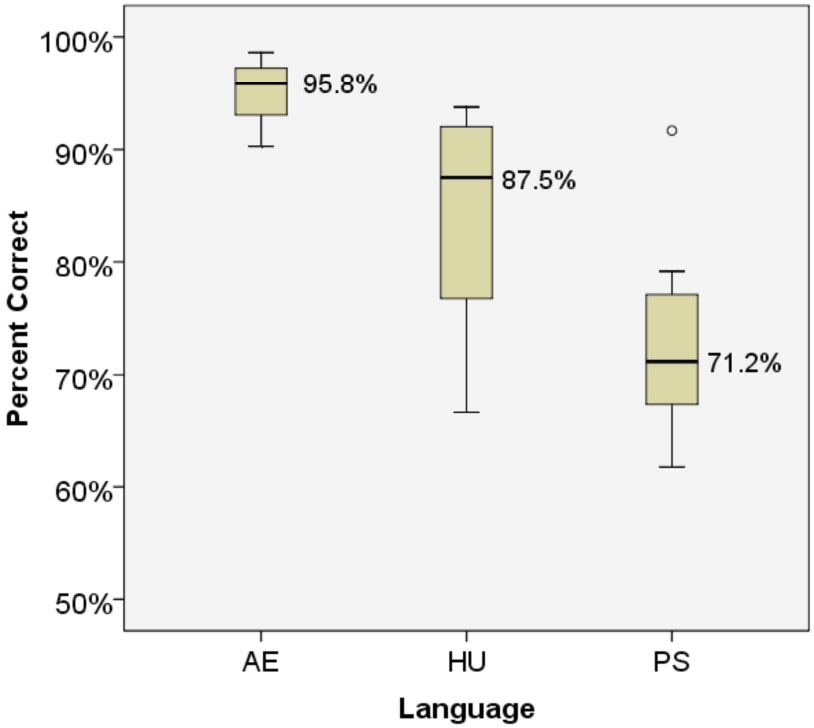
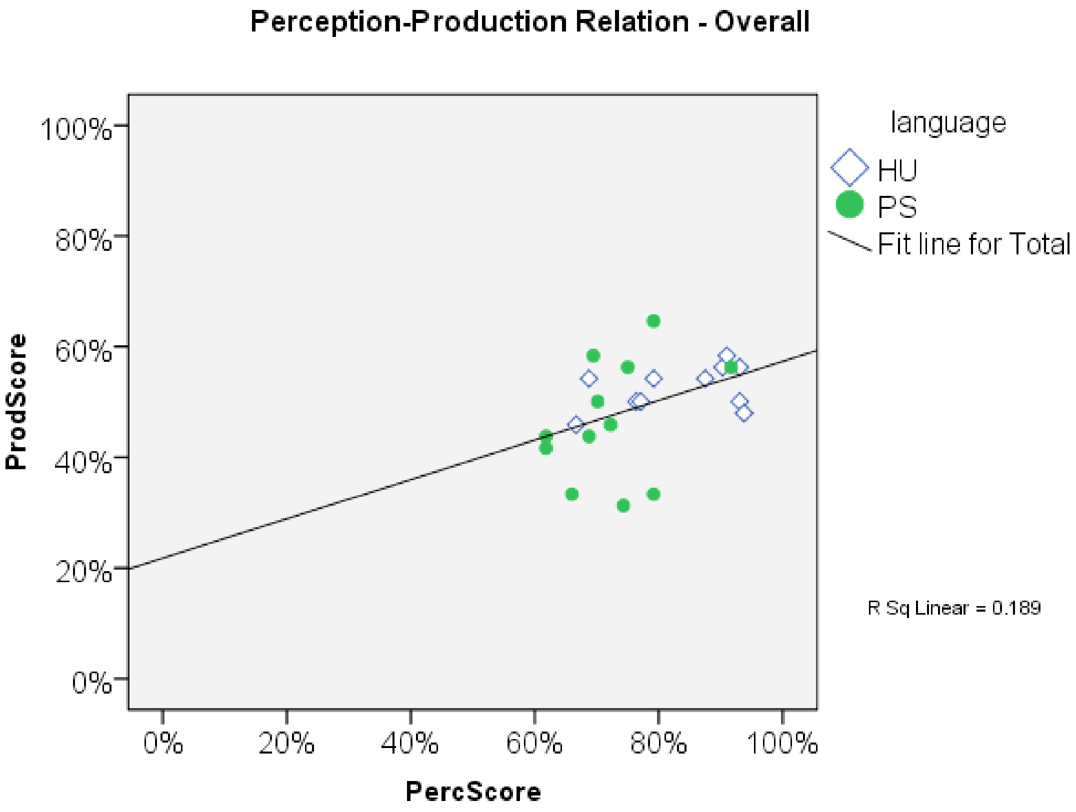


Figure 3.4

Scatterplot illustrating the relationship between non-native perception and production scores. The perception score was based on the overall word-final percent correct in the four-choice task. The production score was based on word-final voicing percentages.



7. Appendices

Appendix A. Language Background Questionnaire (All Participants)

Please complete this questionnaire to the best of your knowledge and add any information you feel might be relevant.
Date: ______________________________ Subject Number: ______________________
Name: ____
Address: ____
E-mail Address: ____
Telephone: (Home) _______________________ (Work/Cell) _________________________
Date of Birth: __________________________ Gender: Male / Female
Birthplace: ____________________________ Native Language: ______________________
Places in which you have lived for more than 1 year:
City/State/Country
__________________________________ from __________ to __________
__________________________________ from __________ to __________
__________________________________ from __________ to __________
__________________________________ from __________ to __________
__________________________________ from __________ to __________
__________________________________ from __________ to __________

Family/Language Background (fill in all that apply):
Mother's/Guardian's birthplace _________________ Language/s spoken ______________________
Father's/Guardian's birthplace _________________ Language/s spoken ______________________
Caretaker's/nanny's birthplace _________________ Language/s spoken ______________________
Grandmother's birthplace _____________________ Language/s spoken ______________________
Grandfather's birthplace ______________________ Language/s spoken ______________________
As a child, what languages were spoken in your home? (for example, by parents, guardians, grandparents, or relatives)
____
What other languages do you speak fluently and understand without effort? ______________________
What languages do you speak & understand with some effort? ______________________________
What language do you consider your "mother tongue"? ___________________________________
What language(s) did you speak/understand as a child (before going to school)?____________________

Language learning settings

Formal Schooling: For each level of schooling, check the language(s) of instruction where applicable. If you check 'Other language', list the relevant language(s) on the right. For each check, write the number of years of instruction in that language in the space provided.

A. Pre-school: ☐ English ☐ Other language ____________________
\# of yrs __________ # of yrs __________
B. Primary/Elementary school: ☐ English ☐ Other language ____________________
\# of yrs __________ # of yrs __________
C. Secondary/High school: ☐ English ☐ Other language ____________________
\# of yrs __________ # of yrs __________
D. College: ☐ English ☐ Other language ____________________
\# of yrs __________ # of yrs __________
E. Graduate/Professional school: ☐ English ☐ Other language ____________________
\# of yrs __________ # of yrs __________
F. Other schooling (conversation classes, etc…): ☐ English ☐ Other language ____________________
\# of yrs __________ # of yrs __________

Other language learning environments/experience *not mentioned above* (including **social interaction**, **work** or other **professional activities**, **extensive travel**, **self-teaching**, etc…). Please specify, and include dates attended.

Language: English
Learning setting 1: ____________________________________ Dates: From ____________ to ______________
Learning setting 2: ____________________________________ Dates: From ____________ to ______________
Learning setting 3: ____________________________________ Dates: From ____________ to ______________
Language: ________________________________
Learning setting 1: ____________________________________ Dates: From ____________ to ______________
Learning setting 2: ____________________________________ Dates: From ____________ to ______________
Learning setting 3: ____________________________________ Dates: From ____________ to ______________
Language: ________________________________
Learning setting 1: ____________________________________ Dates: From ____________ to ______________
Learning setting 2: ____________________________________ Dates: From ____________ to ______________
Learning setting 3: ____________________________________ Dates: From ____________ to _______________

Appendix B. Language Background Questionnaire – Additional Page for Non-native Participants

Language Use

Please check the number that corresponds with the amount that you generally use in the following situations. Try to base your estimate on your use of your native language or English over the past 5 years in the USA or other English-speaking country.
If you use a language *other* than your native language or English in any of these environments, please make a note of that in the space provided below the table.

	1	2	3	4	5	6	7	Not applicable
	Your native language almost all the time (90-100% native language)	Your native language usually with a little English (75-89% native language)	Your native language with some English (60-74% native language)	Your native language as much as English (40-59% native language)	English usually with some of your native language (25-39% native language)	English usually with a little of your native language (10-24% native language)	English almost all the time (0-9% native language)	
At home:								
Listening and speaking								
Reading and writing								
Listening/watching TV, radio, internet, movies, etc…								
At school/work:								
Listening and speaking								
Reading and writing								
In your community:								
Listening and speaking								
Reading and writing								

If you use languages other than your native language or English, please explain the circumstances in which you use those language specifically.
Language ____________________ Circumstance ____
Language ____________________ Circumstance ____
Language ____________________ Circumstance ____

Appendix C. Target and Practice Stimuli for Perception Task

Each stimulus was embedded in the carrier phrase *I met* ____________ *today.*

	Word-final seq.	First Name	Last Name
Stop-stop	V_1K/V_1G (vel)	Mack	Pillman
	V_1K/V_1G (vel)	Mag	Billman
	V_2P/V_2B (lab)	Depp	Geller
	V_2P/V_2B (lab)	Deb	Keller
	V_1K/V_1G (vel)	Mack	Billman
	V_1K/V_1G (vel)	Mag	Pillman
	V_2P/V_2B (lab)	Depp	Keller
	V_2P/V_2B (lab)	Deb	Geller
Stop-Fricative	V_1K/V_1G (vel)	Mack	Zimmer
	V_1K/V_1G (vel)	Mag	Zimmer
	V_2P/V_2B (lab)	Depp	Zimmer
	V_2P/V_2B (lab)	Deb	Zimmer
	V_1K/V_1G (vel)	Mack	Simmer
	V_1K/V_1G (vel)	Mag	Simmer
	V_2P/V_2B (lab)	Depp	Simmer
	V_2P/V_2B (lab)	Deb	Simmer
Fricative-Stop	V_2S/V_2Z (alv)	Kass	Pillman
	V_2S/V_2Z (alv)	Kaz	Billman
	V_1S/V_1Z (alv)	Jess	Keller
	V_1S/V_1Z (alv)	Jez	Geller
	V_2S/V_2Z (alv)	Kass	Billman
	V_2S/V_2Z (alv)	Kaz	Pillman
	V_1S/V_1Z (alv)	Jess	Geller
	V_1S/V_1Z (alv)	Jez	Keller

Filler/Practice stimuli:

Each stimulus was embedded in the carrier phrase *I met* ____________ *today.*

First Name	Last Name
Belle	Allen
Ben	Ullman
Belle	Ullman
Ben	Allen
Jen	Wyman
Jem	Ryman
Jen	Ryman
Jem	Wyman
Sal	Henley
Sam	Fenley
Sam	Henley
Sal	Fenley
Belle	Henley
Ben	Fenley
Belle	Henley
Ben	Fenley
Jen	Allen
Jem	Ullman
Jen	Ullman
Jem	Allen
Sal	Wyman
Sam	Ryman
Sam	Ryman
Sal	Wyman

Appendix D. Sample Page from Protocol

1. I met Jess Geller today.
2. I met Depp Simmer today.
3. I met Sal Henley today.
4. I met Kass Pillman today.
5. I met Mack Zimmer today.
6. I met Belle Allen today.
7. I met Jez Keller today.
8. I met Deb Zimmer today.
9. I met Ben Fenley today.
10. I met Mag Billman today.
11. I met Depp Geller today.
12. I met Jenn Wyman today.
13. I met Kaz Billman today.
14. I met Mag Simmer today.
15. I met Sam Ryman today.
16. I met Deb Keller today.
17. I met Jem Ullman today.
18. I met Mack Pillman today.
19. I met Jess Geller today.

Appendix E. Full text of instructions to the perception experiment

Instructions presented at the beginning the practice session:

Screen 1:

You are going to listen to a series of English names embedded in the sentence:

"I met ___________ today"

Your task is to identify what you heard out of a set of choices presented on the computer screen.

For example, after listening to the phrase:
"I met Ben Henley today"

You will be given the following choices:

Ben Henley	Belle Henley
Ben Fenley	Belle Fenley

And you will click on the one that matches what you heard, that is, the top left item in this example.

Click the space bar to continue

Screen 2:

Some of the names you hear will be full names containing a first and last name, as in

"I met Ben Henley today"

Other names will only contain either a first or last name, as in

"I met Ben today"
or
"I met Henley today"

You may notice that some of the names sound odd or chopped off, and you may find some of the items quite challenging. For each item, just answer as best you can. If you are not sure of your answer, guess. People tend to do better on this task than they think.

Press the space bar to continue.

Screen 3:

You are about to begin the practice session. You will be presented with 32 sentences, presented in 2 blocks of 16 sentences each.

If you have any questions, please ask the experimenter now.

Otherwise, you may put the headphones on and press the space bar to begin.

Instructions presented before the experimental trials:

You are now ready to begin the experiment.

There will be 9 blocks total, each containing 48 sentences.

Remember, if you are unsure of an answer, just guess.

You may take a break at any time by waiting to press "Click to Continue" between sentences.

Press the space bar when you are ready to continue.

Appendix F. Acoustic Measurements of Perception Stimuli

Summary of acoustic measurement data for fricative–stop perception stimuli.

Voicing Sequence	Name	Vowel Duration (ms)		Fricative Duration (ms)		% Fricative Voicing		VOT(ms)	
		Median	Range	Median	Range	Median	Range	Median	Range
Vd Vd	Jez Geller	140	134–165	59	55–66	100	100–100	-48	-51 – -44
	Kaz Billman	137	118–156	60	54–75	100	100–100	-61	-73 – -52
Vd Vl	Jez Keller	137	128–148	65	60–71	39.8	33.0–56.7	61	45–74
	Kaz Pillman	155	137–171	62	50–68	37.3	23.1–46.1	42	38–56
Vl Vd	Jess Geller	119	113–132	84	69–96	10.4	4.5–16.2	11*	8–12*
	Kass Billman	138	117–165	74	65–84	7.0	4.7–10.0	-92	-107 – -33
Vl Vl	Jess Keller	124	115–136	83	78–90	7.7	6.2–9.5	57	45–66
	Kass Pillman	135	114–149	70	61–72	8.6	0–10.2	55	45–73

* Median and range based on 5 tokens. The 6th token had negative VOT (-72 ms).

Summary of acoustic measurement data for stop–stop perception stimuli.

Voicing Sequence	Name	Vowel Duration (ms) Median	Vowel Duration (ms) Range	% Closure Voicing Median	% Closure Voicing Range	VOT (ms) Median	VOT (ms) Range
Vd Vd	Deb Geller	105	92–107	100	100–100	-48	-61 – -38
	Mag Billman	132	119–143	100	100–100	-61	-95 – -43
Vd Vl	Deb Keller	102	95–109	100	100–100	61	51–83
	Mag Pillman	130	120–146	100	70.5–100	53	32–69
Vl Vd	Mack Billman	122	114–125	0	0–0	8	8–10
	Depp Geller	88	78–92	24.3	12.6–45.2	14	11–17
Vl Vl	Depp Keller	81	76–90	20.6	0–37.2	59	40–94
	Mack Pillman	125	117–134	0	0–0	50	41–55

Summary of acoustic measurement data for stop–fricative perception stimuli.

Voicing Sequence	Name	Vowel Duration (ms)		% Closure Voicing		Fricative Duration (ms)		% Fricative Voicing	
		Median	Range	Median	Range	Median	Range	Median	Range
Vd Vd	Deb Zimmer	100	89–114	100	100–100	78	68–89	100	58.2–100
	Mag Zimmer	125	113–145	100	100–100	69	54–87	100	100–100
Vd Vl	Deb Simmer	94	91–108	100	100–100	122	113–150	0	0–8.1
	Mag Simmer	133	124–153	100	100–100	112	103–128	0	0–15.0
Vl Vd	Depp Simmer	83	76–88	33.7	21.1–51.5	134	126–147	0	0–0
	Mack Zimmer	117	108–124	0	0–0	100	78–122	82.8	12.5–100
Vl Vl	Depp Zimmer	84	76–90	21.8	11.7–67.8	106	81–137	100	40.5–100
	Mack Simmer	123	113–140	4.2	0–20.9	113	109–136	0	0–0

Appendix G. Stimuli for Production Task

Target Stimuli

First names are near minimal pairs; all last names are disyllabic
Carrier sentence: *I met first name, last name today*.

	Vowel	Word-final p.o.a	Word-final voice	Word-initial p.o.a.	Word-initial voice	First name	Last Name
Stop-Stop	ʌ	Vel	-	Lab	-	Buck	Potter
	ʌ	Vel	+	Lab	+	Jug	Barker
	ɪ	Lab	-	Vel	-	Kipp	Kenner
	ɪ	Lab	+	Vel	+	Gib	Guffman
	ʌ	Vel	-	Lab	+	Buck	Barker
	ʌ	Vel	+	Lab	-	Jug	Potter
	ɪ	Lab	-	Vel	+	Kipp	Guffman
	ɪ	Lab	+	Vel	-	Gib	Kenner
Stop-Fricative	ʌ	Vel	-	Alv	+	Buck	Zelder
	ʌ	Vel	+	Alv	+	Jug	Zelder
	ɪ	Lab	-	Alv	+	Kipp	Zelder
	ɪ	Lab	+	Alv	+	Gib	Zelder
	ʌ	Vel	-	Alv	-	Buck	Simpson
	ʌ	Vel	+	Alv	-	Jug	Simpson
	ɪ	Lab	-	Alv	-	Kipp	Simpson
	ɪ	Lab	+	Alv	-	Gib	Simpson
Fricative-Stop	ʌ	Alv	-	Lab	-	Gus	Potter
	ʌ	Alv	+	Lab	+	Buzz	Barker
	ɪ	Alv	-	Vel	-	Siss	Kenner
	ɪ	Alv	+	Vel	+	Liz	Guffman
	ʌ	Alv	-	Lab	+	Gus	Barker
	ʌ	Alv	+	Lab	-	Buzz	Potter
	ɪ	Alv	-	Vel	+	Siss	Guffman
	ɪ	Alv	+	Vel	-	Liz	Kenner

Filler Stimuli

	Vowel	Wd-final p.o.a	Word-final voice	Wd-initial p.o.a.	Wd-initial voice	Noun	Verb
Stop-stop	æ	Vel	-	Lab	-	Sack	Pours
	æ	Vel	+	Lab	-	Bag	Puts
	ʌ	Lab	-	Vel	+	Pup	Gives
	ʌ	Lab	+	Vel	+	Tub	Gets
	ɛ	Vel	-	Lab	+	Deck	Becomes
	ɛ	Vel	+	Lab	+	Peg	Belongs
	ɪ	Lab	-	Vel	-	Dip	Comes
	ɪ	Lab	+	Vel	-	Bib	Keeps
Stop-Fricative	ʌ	Vel	-	Alv	-	Duck	Sits
	ʌ	Vel	+	Alv	-	Mug	Soaks
	ɛ	Vel	-	Alv	+	Neck	Zips
	ɛ	Vel	+	Alv	+	Keg	Zig-zags
	æ	Lab	-	Alv	-	Cap	Sells
	æ	Lab	+	Alv	-	Tab	Seems
	ɪ	Lab	-	Alv	+	Tip	Zooms
	ɪ	Lab	+	Alv	+	Fib	Zaps
Fricative-Stop	ʌ	Alv	-	Lab	-	Bus	Parks
	ʌ	Alv	+	Lab	-	Fuzz	Piles
	æ	Alv	-	Lab	+	Gas	Bothers
	æ	Alv	+	Lab	+	Jazz	Begins
	ʌ	Alv	-	Vel	-	Fuss	Kills
	ʌ	Alv	+	Vel	-	Buzz	Calls
	ɛ	Alv	-	Vel	+	Mess	Gains
	ɛ	Alv	+	Vel	+	Fez	Goes
Stop-Sonorant	ʌ	Vel	+	Son		Bug	Likes
	ʌ	Vel	-	Son		Luck	Lets
	æ	Lab	-	Son		Nap	Makes
	æ	Lab	+	Son		Cab	Meets
Fricative-Sonorant	aɪ	Alv	-	Son		Mice	Nestle
	aɪ	Alv	+	Son		Size	Needs
VelStop#	æ	Vel	-			Back	
	æ	Vel	+			Tag	
	ɪ	Vel	+			Pig	
	ɪ	Vel	-			Chick	
LabStop#	ʌ	Lab	-			Cup	
	ʌ	Lab	+			Pub	
Fricative#	ɪ	Alv	-			Kiss	
	ɪ	Alv	+			Liz	

Sentences:
Her ba**g p**uts extra strain on her shoulder.
The pu**p g**ives her a big ki**ss**.
The di**p c**omes with potato chips.
The du**ck s**its down next to her baby chi**ck**.
The tu**b g**ets full very quickly.
The ti**p z**ooms across the counter.
The bu**s p**arks outside next to the pu**b**.
The ja**zz b**egins to play as the diners eat their first course.
His ca**b m**eets him at the building in the ba**ck**
The bu**g l**ikes to fly around the pi**g**.
Her lu**ck l**ets her travel around the world.
The bu**zz c**alls the workers back to the office.
The fu**zz p**iles up as she dries the clothes.
The pe**g b**elongs to the dining room table.
The mi**ce n**estle together in the space behind the wall.
Her na**p m**akes her feel even sleepier than she was before.
The fe**z g**oes on top of his head.
The si**ze ne**eds to be altered by Li**z**.
The ca**p s**ells for twenty dollars, according to the price ta**g.**
The bi**b k**eeps the baby clean after drinking from the cu**p**.
The ne**ck z**ips up too tightly on Peter’s jacket.
The fu**ss k**ills Mary’s good mood.
The ga**s b**others Linda because it is too expensive.
The ta**b s**eems lower than it should be.
The fi**b z**aps Ann of all her energy.
The mu**g s**oaks in the kitchen sink.
The ke**g z**igzags across the yard after falling from the truck.
Her de**ck b**ecomes very festive when they decorate it for the holidays.
The sa**ck p**ours out ten pounds of flour.
The me**ss g**ains more clutter when the children arrive.

Appendix H:
Median percent word-final perception accuracy across voicing contexts (four-choice task)

		American English		Hungarian		Polish	
		Median%	Range	Median%	Range	Median%	Range
Fric-Stop	Vd Vd	**100**	91.7-100	**100**	66.7-100	**95.8**	33.0-100
	Vd Vls	**95.8**	83.3-100	**91.7**	50.0-100	**70.8***	33.0-100
	Vls Vd	**95.8**	58.3-100	**83.3***	50.0-100	**75.0***	33.0-91.7
	Vls Vls	**83.3**	66.7-100	**91.7**	41.7-100	**83.3**	58.3-91.7
Stop-Stop	Vd Vd	**100**	91.7-100	**100**	50.0-100	**62.5***	33.3-100
	Vd Vls	**100**	91.7-100	**91.7**	16.7-100	**62.5***	33.3-100
	Vls Vd	**95.8**	75.0-100	**83.3***	50.0-100	**75.0***	16.7-100
	Vls Vls	**100**	100-100	**91.7***	58.3-100	**91.7***	50.0-100
Stop-Fric	Vd Vd	**100**	75.0-100	**100**	66.7-100	**87.5**	50.0-100
	Vd Vls	**91.7**	58.3-100	**75.0***	16.7-100	**50.0***	25.0-83.3
	Vls Vd	**100**	83.3-100	**83.3***	25.0-100	**66.7***	41.7-91.7
	Vls Vls	**100**	91.7-100	**75.0***	41.7-100	**87.5***	41.7-100

*An asterisk indicates that the score was significantly lower ($p < .05$, one-tailed) than the AE score for the same context

Appendix I. Verification of Acoustic Measurements

Average difference between first and second measurements of 10% of production tokens

	Vowel Duration (ms)		Fricative Duration (ms)		% Fricative Voicing		% Closure Voicing		Negative VOT (ms)		Positive VOT (ms)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Fric-stop	2.5	3.0	2.2	2.0	2.1	3.0	-	-	-	-	1.1	1.4
Stop-stop	1.9	2.2	-	-	-	-	2.9	5.1	13.1	2.9	2.4	5.0
Stop-fric	2.0	2.0	2.4	4.0	.5	2.0	1.5	3.1	-	-	-	-

Appendix J. Acoustic Measurement Data for Production Task

Table 1

Summary of acoustic measurement data for fricative-stop productions: AE participants (number of tokens analyzed = 16 per type)

		Vowel Duration (ms)		Fricative Duration (ms)		% Fricative Voicing		VOT (ms)	
Voicing Sequence	Name	Median	I-Q Range	Median	I-Q Range	Median	I-Q Range	Median*	I-Q Range
Vd Vd	Buzz Barker	134	120–142	61	53–69	100	100–100	Neg(4): -92 Pos(12): 10	-98 – -74 7–11
	Liz Guffman	113	107–124	78	69–91	56	41–100	Neg(4): -49 Pos(12): 17	-53 – -32 15–22
Vd Vl	Buzz Potter	138	131–157	59	51–81	46	32–55	Pos(16): 47	28–69
	Liz Kenner	108	105–122	81	75–95	31	23–37	Pos(16): 64	50–79
Vl Vd	Gus Barker	109	102–119	81	73–93	9	7–11	Neg(3): -64 Pos(13): 9	-93 – -61 7–10
	Siss Guffman	66	49–71	97	90–114	13	10–20	Neg(1): -59 Pos(15): 19	-59 17–23
Vl Vl	Gus Potter	105	99–110	76	71–95	7	3–10	Pos(16): 53	23–66
	Siss Kenner	72	68–78	106	101–117	10	7–12	Pos(16): 63	57–72

*Number in parentheses indicates number of tokens

Table 2

Summary of acoustic measurement data for fricative-stop productions: HU participants (number of tokens analyzed = 22 per type)

Voicing Sequence	Name	Vowel Duration (ms)		Fricative Duration (ms)		% Fricative Voicing		VOT (ms)	
		Median	I-Q Range	Median	I-Q Range	Median	I-Q Range	Median*	I-Q Range
Vd Vd	Buzz Barker	147	130–156	69	59–92	53	19–100	Neg(9): -95 Pos(13): 7	-101–-49 7–12
	Liz Guffman	103	82–121	85	74–94	37	19–100	Neg(7): -52 Pos(15): 23	-64–-46 20–25
Vd Vl	Buzz Potter	139	124–152	79	66–97	6	0–9	Pos(22): 17	11–36
	Liz Kenner	105	90–118	110	101–116	8	6–10	Pos(22): 36	26–50
Vl Vd	Gus Barker	137	124–155	85	69–101	13	5–28	Neg(10): -53 Pos(12): 8	-93–-24 7–10
	Siss Guffman	96	87–108	98	86–106	17	9–31	Neg(3): -50 Pos(19): 21	-78–-50 18–23
Vl Vl	Gus Potter	137	115–149	93	81–109	2	0–7	Pos(22): 15	10–27
	Siss Kenner	86	73–95	127	106 –141	7	6–10	Pos(22): 33	28–47

*Number in parentheses indicates number of tokens

Table 3

Summary of acoustic measurement data for fricative-stop productions: PS participants (number of tokens analyzed = 24 per type)

Voicing Sequence	Name	Vowel Duration (ms)		Fricative Duration (ms)		% Fricative Voicing		VOT (ms)	
		Median	I-Q Range	Median	I-Q Range	Median	I-Q Range	Median*	I-Q Range
Vd Vd	Buzz Barker	132	104–159	75	58 –102	68	23–100	Neg(19): -86 Pos(5): 9	-96 – -79 7 –13
	Liz Guffman	100	76–120	100	86–112	32	25–90	Neg(11): -56 Pos(13): 27	-69 – -51 15–24
Vd Vl	Buzz Potter	142	122–162	94	82–112	10	6–17	Pos(24): 22	11–34
	Liz Kenner	96	79–114	125	103–144	14	9–19	Pos(24): 39	33–55
Vl Vd	Gus Barker	130	97–158	98	70–126	10	5–35	Neg(17): -83 Pos(7): 11	-92 – -70 8–12
	Siss Guffman	71	62–85	95	77–120	28	13–49	Neg(14): -61 Pos(10): 21	-70 – -47 18–28
Vl Vl	Gus Potter	119	103 –135	103	78–127	6	0–8	Pos(24): 20	11–42
	Siss Kenner	75	59–90	135	118–154	11	8–14	Pos(24): 38	31–59

*Number in parentheses indicates number of tokens

Table 4

Summary of acoustic measurement data for stop-stop productions: AE participants (number of tokens analyzed = 16 per type)

Voicing Sequence	Name	Vowel Duration (ms)		% Closure Voicing		VOT (ms)	
		Median	I-Q Range	Median*	I-Q Range	Median*	I-Q Range
Vd Vd	Jug Barker	124	116–140	Burst(5): 100	89–100	Neg(12): -97	-114 – -69
				NoBurst(11): 100	40–100	Pos(4): 8	6–10
	Gib Guffman	93	76–104	Burst(15): 100	100–100	Neg(6): -53	-59 – -42
				NoBurst(1): 55	55	Pos(10): 20	15–28
Vd Vl	Jug Potter	114	106–129	Burst(5): 100	75–100	Pos(16): 48	33–73
				NoBurst(11): 49	21–56		
	Gib Kenner	90	76–103	Burst(16): 100	100–100	Pos(16): 65	56–76
Vl Vd	Buck Barker	93	87–106	Burst(9): 0	0–18	Neg(1): -179	-179
				NoBurst(7): 0	0–3	Pos(15): 7	6–10
	Kip Guffman	64	51–74	Burst(15): 35	27–52	Neg(1): -56	-56
				NoBurst(1): 21	21	Pos(15): 19	14–26
Vl Vl	Buck Potter	94	79–103	Burst(10): 0	0–0	Pos(16): 52	36–61
				NoBurst(6): 0	0–1		
	Kip Kenner	60	49–71	Burst(15): 29	21–42	Pos(16): 66	57–75
				NoBurst(1): 13	13		

*Number in parentheses indicates number of tokens

Table 5

Summary of acoustic measurement data for stop-stop productions: HU participants (number of tokens analyzed = 22 per type)

		Vowel Duration (ms)		% Closure Voicing		VOT (ms)	
Voicing Sequence	Name	Median	I-Q Range	Median*	I-Q Range	Median*	I-Q Range
Vd Vd	Jug Barker	131	117–146	Burst(3): 100	100–100	Neg(15): -119	-126 – -103
				NoBurst(19): 90	39–100	Pos(7): 10	7–13
	Gib Guffman	94	75–108	Burst(21): 100	54–100	Neg(8): -66	-80 – -41
				NoBurst(1): 54	54	Pos(14): 22	18–28
Vd Vl	Jug Potter	137	125–144	Burst(15): 0	0–0	Pos(22): 21	12–31
				NoBurst(7): 0	0–0		
	Gib Kenner	91	78–106	Burst(20): 37	20–54	Pos(22): 47	44–60
				NoBurst(2): 11	10–13		
Vl Vd	Buck Barker	123	106–138	Burst(13): 14	0–64	Neg(14): -102	-130 – -79
				NoBurst(9): 100	20–100	Pos(8): 11	8–14
	Kip Guffman	64	61–77	Burst(20): 43	33–78	Pos(22): 50	43–65
				NoBurst(2): 55	50–59		
Vl Vl	Buck Potter	127	109–140	Burst(18): 0	0–0	Pos(22): 16	13–39
				NoBurst(4): 0	0–0		
	Kip Kenner	67	51–80	Burst(22): 27	16–38	Neg(5): -39	-85 – -31
						Pos(17): 25	22–27

*Number in parentheses indicates number of tokens

Table 6

Summary of acoustic measurement data for stop-stop productions: PS participants (Number of tokens analyzed = 24)

		Vowel Duration (ms)		% Closure Voicing		VOT (ms)	
Voicing Sequence	Name	Median	I-Q Range	Median*	I-Q Range	Median*	I-Q Range
Vd Vd	Jug Barker	124	109–156	Burst(20): 59	3–100	Neg(21): -89	-107 – -82
				NoBurst(4): 100	100–100	Pos(3): 11	10–12
	Gib Guffman	86	64–106	Burst(20): 86	41–100	Neg(11): -72	-93 – -57
				NoBurst(4): 87	52–100	Pos(13): 24	18–34
Vd Vl	Jug Potter	121	109–143	Burst(21): 13	0–29	Pos(24): 20	14–39
				NoBurst(3): 0	0–0		
	Gib Kenner	95	65–105	Burst(24): 41	26–55	Pos(24): 60	44–76
Vl Vd	Buck Barker	115	99–128	Burst(21): 27	0–70	Neg17): -92	-106 – - 78
				NoBurst(3): 100	100–100	Pos(7): 10	7–13
	Kip Guffman	58	41–73	Burst(22): 48	31–85	Neg(15): -65	-73 – -56
				NoBurst(2): 60	19–100	Pos(9): 26	22–27
Vl Vl	Buck Potter	122	105–139	Burst(22): 0	0–10	Pos(24): 22	18–41
				NoBurst(2): 0	0–0		
	Kip Kenner	46	38–64	Burst(24): 31	25–46	Pos(24): 53	44–69

*Number in parentheses indicates number of tokens

Table 7

Summary of acoustic measurement data for stop-fricative productions: AE participants (number of tokens analyzed = 16 per type)

Voicing Sequence	Name	Vowel Duration (ms)		% Closure Voicing		Fricative Duration		% Fricative Voicing	
		Median	I-Q Range	Median	I-Q Range	Median	I-Q Range	Median	I-Q Range
Vd Vd	Jug Zelder	131	118–139	100	100–100	94	79–101	100	100–100
	Gib Zelder	98	80–124	100	100–100	98	82–114	100	100–100
Vd Vl	Jug Simpson	120	112–137	100	100–100	134	113–150	0	0–0
	Gib Simpson	85	73–100	100	89–100	134	116–155	0	0 –10
Vl Vd	Buck Zelder	99	92–112	6	0–38	107	90–122	56	0–100
	Kip Zelder	63	53–73	61	22–100	108	96–119	100	48–100
Vl Vl	Buck Simpson	97	84–104	0	0–15	133	110–146	0	0–0
	Kip Simpson	59	48–72	31	15–41	135	110–147	0	0–0

Table 8

Summary of acoustic measurement data for stop-fricative productions: HU Participants (number of tokens analyzed = 22 per type).

Voicing Sequence	Name	Vowel Duration (ms)		% Closure Voicing		Fricative Duration		% Fricative Voicing	
		Median	I-Q Range	Median	I-Q Range	Median	I-Q Range	Median	I-Q Range
Vd Vd	Jug Zelder	130	110–146	100	100–100	103	92–116	79	13–100
	Gib Zelder	88	74–103	100	94–100	113	92–125	41	7–100
Vd Vl	Jug Simpson	124	116–138	0	0–22	134	123–161	0	0–0
	Gib Simpson	94	73–109	45	21–66	142	122–64	0	0–0
Vl Vd	Buck Zelder	128	113–135	48	0–100	113	104–130	32	0–81
	Kip Zelder	69	61–79	77	39–100	115	93–137	25	0–39
Vl Vl	Buck Simpson	112	105–129	0	0–0	137	115–150	0	0–0
	Kip Simpson	69	59 –77	30	17–40	135	127–155	0	0–0

Table 9

Summary of acoustic measurement data for stop-fricative productions: PS participants (number of tokens analyzed = 24 per type)

		Vowel Duration (ms)		% Closure Voicing		Fricative Duration		% Fricative Voicing	
Voicing Sequence	Name	Median	I-Q Range	Median	I-Q Range	Median	I-Q Range	Median	I-Q Range
Vd Vd	Jug Zelder	122	103–150	100	61–100	114	93–129	100	34–100
	Gib Zelder	79	62–105	100	57–100	120	100–132	100	28–100
Vd Vl	Jug Simpson	134	108–155	22	2–50	146	133–163	0	0–0
	Gib Simpson	78	58–93	50	30–71	143	135–166	0	0–0
Vl Vd	Buck Zelder	109	95 –135	35	0–100	113	93–134	85	0–100
	Kip Zelder	60	49–73	62	37–100	131	103–141	100	13–100
Vl Vl	Buck Simpson	101	97–134	0	0–19	147	136 –178	0	0–0
	Kip Simpson	47	32–67	37	25–50	149	132–168	0	0–0

8. References

Abramson, A. S. & Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics, 1*, 1-8.

Altenberg, E. P., & Vago, R. M. (1983). Theoretical implications of an error analysis of second language phonology production. *Language Learning, 33*, 427-447.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research* (pp. 171-204). Timonium, MD: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O. -S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production.* Amsterdam: John Benjamins.

Boersma, P. & Weeknik, D. (2007). Praat: Doing phonetics by computer (Version 5.0.06) [Computer Program]. Retrieved October 8, 2007, from http://www.praat.org/

Broersma, M. (2005). Perception of unfamiliar contrasts in familiar positions. *Journal of the Acoustical Society of America, 117*(6)*,* 3890-3901.

Darcy, I., Peperkamp, S., & Dupoux, E. (2005). Bilinguals play by the rules. Perceptual compensation for assimilation in late L2-learners. *Papers in Laboratory Phonology, 9*.

Dupoux, E., Kakehi K., Hirose, Y., Pallier, C., and Mehler, J. (1999). Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25*(6), 1568-1578.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, *27*, 315-330.

Eckman, F. R. (2008). Typological markedness and second language phonology. In Hansen Edwards, J. G. and Zampini, M. L. (eds.) *Phonology and Second Language Acquisition* (pp. 95-115). Amsterdam and Philadelphia: John Benjamins.

Fais, L., Kajikawa, S., Werker, J., & Amano, S. (2005). Japanese Listeners' Perceptions of Phonotactic Violations. *Langauge and Speech, 48*(2), 185-201.

Flege, J. E. (1989). The perception of /t/ and /d/ by native and Chinese listeners. *Journal of the Acoustical Society of America*, *84*, 1639-1652.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and Problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233-277). Timonium, MD: York Press.

Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica, 43*(4),155-171.

Flege J. E., & Hillenbrand J. (1986). Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. *Journal of the Acoustical Society of America, 79*(2), 508-517.

Flege, J. E., Munro, M. J., and Fox, R. A. (1994). Auditory and categorial effects on cross-language vowel perception. *Journal of the Acoustical Society of America, 95*, 3623-3641.

Flege, J. E., & Wang, C. (1989). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/ - /d/ contrast. *Journal of Phonetics, 17*, 299-315.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 144-158.

Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological interference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 380-396.

Gaskell, M. G., & Marslen-Wilson, W. D. (2001). Lexical ambiguity and spoken word recognition: bridging the gap. *Journal of Memory and Language*, *44*, 325-349.

Gosy, M. (2001). The VOT of the Hungarian voiceless plosives in words and in spontaneous speech. *International Journal of Speech Technology, 4*(1), 75-85.

Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language, 45*, 133-159.

Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance, 28*, 163-179.

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception and Psychophysics*, *65*, 575-590.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America, 107*, 2711-2724.

Gussman, E. (2007). The Phonology of Polish. Oxford University Press.

Hallé, P. A., Segui, J., Frauenfelder, U., and Meunier, C. (1998). Processing of illegal consonant clusters: a case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 592-608.

Jansen, W. (2004). *Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch.* Ph.D. thesis, Rijksuniversiteit, Groningen.

Jansen, W. (2007). English fricatives and the typology of regressive voicing assimilation. *Language Sciences*, *29*(2-3)*,* 270-293.

Jassem, W., & Richter, L. (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics*, *17*, 317-325.

Keating, P. A. (1984). Phonetic and Phonological Representation of Stop Consonant Voicing. *Language, 60*, 286-319.

Keating, P. A., Mikoś, M. J., & Ganong, W. F. (1981). A cross-language study of voice onset time in the perception of initial stop voicing. *The Journal of the Acoustical Society of America*, *70*(5), 1261-1271.

Ladefoged, P. (1993). *A course in phonetics* (3rd. ed.). New York: Harcourt Brace Jovanovich.

Lado, R. (1957). *Linguistics across cultures.* Ann Arbor: University of Michigan Press.

Levy, E. S. (2009). Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *Journal of the Acoustical Society of America*, *25*, 1138-1152.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384-422.

MacKay, I. R. A., Flege, J. E., Piske, T., Schirru, C. (2001). Category restructuring during second-language speech acquisition. *Journal of the Acoustical Society of America*, *110*(1), 516-528.

Otake, T., Yoneyama, K., Cutler, A., and van der Lugt, A. (1996). The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America, 100*(6), 3831-3842.

Petrova, O., Plapp, R., Ringen, C., & Szentgyorgyi, S. (2006). Voice and aspiration: evidence from Russian, Hungarian, German, Swedish, and Turkish. *The Linguistic Review, 23*(1)*,* 1-35.

Pikser, R. (2003). *Perception of word-final /t/, /k/, /s/ and /z/ by beginning Spanish speaking learners of English*. Unpublished master's thesis, The Graduate Center of the City University of New York, New York, NY.

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *The Journal of the Acoustical Society of America*, *51*(4), 1296-1303.

Raphael, L. J. (1981). Durations and contexts as cues to word-final cognate opposition in English. *Phonetica*, *38*, 126-147.

Rubach, J. (1984). Rule typology and phonological interference. In S. Eliasson (Ed.) *Theoretical Issues in Contrastive Phonology: Studies in descriptive linguistics* (pp. 539-550). Heidelberg: Julius Groos.

Rubach, J. (1996). Nonsyllabic analysis of voice assimilation in Polish. *Linguistic Inquiry*, *27*(1), 69-110.

Slowiaczek, L. M., & Dinnesen, D. A. (1985). On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics*, *13*, 325-341.

Strange, W. Akahane-Yamada, R., Kubo, R. Trent, S. A., and Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *Journal of the Acoustical Society of America*, *109*, 1691-1704.

Strange, W. and Shafer, V. L. (2008). Speech perception in second language learners: the re-education of selective perception. In Hansen Edwards, J. G. and Zampini, M. L. (eds.) *Phonology and Second Language Acquisition* (pp. 153-191). Amsterdam and Philadelphia: John Benjamins.

Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., & Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America, 91*(5)*,* 2979-3000.

Tagliaferri, B. (2007). Paradigm ID [Compter Software]. New York, NY: City University of New York Graduate Center, Speech and Hearing Sciences.

Vago, R. M. (1980). The sound pattern of Hungarian. Georgetown University Press, Washington.

Williams, L. (1977). The voicing contrast in Spanish. *Journal of Phonetics, 5*, 169-184.