

March 2015

A Heart Thing to Hear But You'll Earn: Processing and Learning about Foreign Accent Features Generated by Phonological Rule Misapplications

Monica Lee Bennett
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/masters_theses_2



Part of the [Cognitive Psychology Commons](#), and the [Psycholinguistics and Neurolinguistics Commons](#)

Recommended Citation

Bennett, Monica Lee, "A Heart Thing to Hear But You'll Earn: Processing and Learning about Foreign Accent Features Generated by Phonological Rule Misapplications" (2015). *Masters Theses*. 138. <https://doi.org/10.7275/6464079> https://scholarworks.umass.edu/masters_theses_2/138

This Open Access Thesis is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

A Heart Thing to Hear But You'll Earn:
Processing and Learning about Foreign Accent Features Generated by Phonological Rule
Misapplications

A Thesis Presented

by

MONICA LEE BENNETT

Submitted to the Graduate School of the
University of Massachusetts, Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

February 2015

Psychological and Brain Sciences

A Heart Thing to Hear But You'll Earn:
Processing and Learning about Foreign Accent Features Generated by Phonological Rule
Misapplications

A Thesis Presented

By

MONICA LEE BENNETT

Approved as to style and content by:

Alexandra Jesse, Chair

Lisa Sanders, Member

John Kingston, Member

Melinda Novak, Department Head
Psychological and Brain Sciences

ACKNOWLEDGEMENTS

Foremost, I would like to thank my advisor, Alexandra Jesse, for her tireless support, compassion, and dedication. Her guidance throughout this process inspired me throughout the most difficult times, and I will be forever grateful that she challenged me to do my best. I would also like to express my gratitude to my other committee members, Lisa Sanders and John Kingston, for their invaluable perspectives and essential advice on this project.

I want to thank my family and friends for always being there for me. Finally, a special thank you to my partner for giving me the strength to see this through.

ABSTRACT

A HEART THING TO HEAR BUT YOU'LL EARN: PROCESSING AND LEARNING ABOUT FOREIGN ACCENT FEATURES GENERATED BY PHONOLOGICAL RULE MISAPPLICATIONS

FEBRUARY 2015

MONICA LEE BENNETT, B.A.SC., MCGILL UNIVERSITY

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Dr. Alexandra Jesse

The present thesis focuses on how native English listeners process phonological rule misapplications in non-native-accented speech. In Experiment 1, we examined whether listeners use information about a speaker's native language to help them understand that speaker's accented English. The test case for this scenario was word-final obstruent devoicing in German and German-accented speech. Results showed that participants did not generalize their knowledge cross-linguistically. In Experiment 2, we used a categorization task and an eye-tracking visual world paradigm to investigate listeners' use of a position-sensitive allophonic alternation, the velarization of /l/, as a word segmentation cue in native English. Participants were able to use velarization as a cue during word segmentation, even though they also showed a later, post-perceptual bias to segment /l/ as word initial. Follow-up experiments will build upon these conclusions using German-accented speech as stimuli, which will have reduced or absent velarization of /l/ in word-final position. In sum, these experiments inform us about the limits of phonological knowledge about foreign-accented speech.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS.....	iii
ABSTRACT.....	iv
LIST OF TABLES.....	vi
LIST OF FIGURES.....	vii
CHAPTER	
I. INTRODUCTION.....	1
II. EXPERIMENT 1.....	11
III. EXPERIMENT 2.....	37
IV. GENERAL DISCUSSION.....	57
APPENDIX: LANGUAGE QUESTIONNAIRE.....	58
REFERENCES.....	60

LIST OF TABLES

Table	Page
1. Examples of test phase primes and targets.....	17
2. Mean percentages of accurate responses and reaction times by group and priming condition.....	28
3. Mean values of acoustic measurements and statistical tests for critical item pairs.....	45

LIST OF FIGURES

Figure	Page
1. Accuracy priming effects in Experiment 1 for G and K words, by exposure group.....	29
2. Mean accuracy scores in Experiment 1 for each group's related versus unrelated prime-target pairs.....	30
3. Reaction time priming effects in Experiment 1 for each type of critical word by exposure group.....	31
4. Fixation proportions to targets, competitors, and distractors in the word-initial /l/ condition.....	50
5. Fixation proportions to targets, competitors, and distractors in the word-final /l/ condition.....	51

CHAPTER I

INTRODUCTION

The English language is estimated to be the second language (L2) of over 430,000,000 speakers around the world (Lewis, Simons, & Fennig, 2013). Encountering foreign-accented English is, accordingly, a common occurrence. Recognizing the speech sounds and words of foreign-accented English can provide a significant challenge for listeners, as foreign-accented English varies from native English in a variety of aspects. Most relevantly for speech recognition, foreign-accented English varies from native English in the realization of segments and prosody. The experiments outlined below aimed to investigate how native English listeners process and learn about foreign-accented features that are created by implementing a phonological rule from the speaker's native language into English (Experiment 1) or by failing to apply a phonological rule in English due to its absence from the speaker's native language (Experiment 2). In Experiment 1, we examined whether listeners use knowledge about a phonological rule in a speaker's native language to process that speaker's accented English. In Experiment 2, we examined effects on native English listeners' segmentation of speech into words when an L2 speaker's accented English is missing a phonological rule of English.

Production in a speaker's L2 often differs systematically from native (L1) speech in the same language. This is especially true for speakers who learned the second language later in life and for speakers who use the language less often in their daily lives. These speakers tend to have accents that are perceived as stronger and less native (Flege, Munro, & MacKay, 1995). Speakers of the same native language show similarities in their L2 speech, which results in what is perceived as a shared "accent". However, even

among speakers of the same native language, there are individual differences in how the accent is realized, determined by a variety of factors, such as proficiency in the L2, age of acquisition, length of residence in the L2 environment, and other types of individual differences (Piske, MacKay, & Flege, 2001). More so, even within the same speaker, application of the accent is not necessarily systematic. While a speaker may at times struggle, for example, with pronouncing a particular non-native phoneme, they may produce it in a more native-like capacity at other times. Given this amount of individual variability among and within speakers with the same foreign accent, evidence suggests that exposure to more than one speaker with the same foreign accent is necessary to learn about the foreign accent such that this knowledge can be generalized to new non-native speakers with the same accent (Bradlow & Bent, 2008).

Foreign-accented speech can vary from native speech such that it can affect all levels of speech recognition. At the phonetic level, phoneme substitutions, additions, or deletions, and subphonemic variations are common features of non-native speech, due to influences from the native language on the production of the second language (Flege, Schirru, & MacKay, 2003). For example, Italian-L1 speakers tend to produce more “Italian-like” [u] in English (i.e., with closer formant values to Italian [u]) than native English speakers’ production of [u] (Busà, 1992, as cited in Munro, Flege, & MacKay, 1996). Another example is that many native speakers of German produce /s/ or /z/ respectively instead of /θ/ and /ð/ when speaking English, as these interdental fricatives do not exist in their phoneme inventory in German (Howell & Dworzynski, 2001). This can lead to ambiguity in processing, as the accented pronunciation of a word matches then less the stored phonological word representation of native listeners and may even

match better the representation of an unintended word (e.g., when “think” is produced as [sɪŋk]). Similarly, phonological rules in the native language can also be applied erroneously to the second language. In German and in Dutch, obstruents in coda position become devoiced, and consequently, native German and Dutch speakers frequently also devoice word-final obstruents when speaking English (Simon, 2008; Smith, Hayes-Harb, Bruss, & Harker, 2009). Native Hungarian speakers sometimes apply a Hungarian assimilation rule in which obstruents are changed to agree in voicing with adjacent obstruents to their English speech, producing sequences like [boylɪŋk pat] for “boiling pot” (Altenberg & Vago, 1983). Not all accent features are segmental, however. Non-native-like prosody also plays a significant role in how “foreign” or strong an accent is perceived to be (Boula de Mareüil, & Vieru-Dimulescu, 2006; Munro, 1995). For example, native French speakers may erroneously place stress on the final syllable of an English word due to the fixed stress pattern of their native tongue, or native Japanese speakers may fail to raise their pitch at the end of a question in English (Wennerstrom, 1994). Foreign accent is therefore shaped by a combination of features varying from native speech.

These differences from native speech can make foreign-accented speech more difficult to understand. Lane’s (1963) seminal experiment was an early demonstration of such difficulties. When listening to speech masked by noise, native English listeners showed significantly lower accuracy when transcribing words produced by foreign-accented speakers with different L1 backgrounds compared to those spoken by native English speakers. Subsequent experiments showed that this decreased transcription accuracy for non-native speech also exists for speech presented without noise, and that it

varies as a function of the foreign-accented speaker's proficiency in the L2 (Rogers, Dalby, & Nishi, 2004). Munro and Derwing (1995) showed that foreign-accented speech also slows down processing. Native English listeners needed more time to determine whether English utterances were true when they were spoken by a native Mandarin speaker than when spoken by a native English speaker. Listeners thus need more time to process foreign-accented speech, and they are less likely to correctly recognize what the speaker is saying.

Despite this increased demand on the language system when recognizing foreign-accented speech, listeners can rapidly adjust to speakers and their accents. L1 listeners initially take more time to process foreign-accented words than native words, but within only minutes of listening, they are able to adapt to the foreign-accented speech (Clarke & Garrett, 2004; Sidaras, Alexander, & Nygaard, 2009). The language system is constantly tasked with adjusting to acoustic variation in L1 speech, given the extent of variation between even individual native speakers of a language. When listeners are exposed to ambiguous sounds whose identity is disambiguated by lexical or visual speech context in native speech, they shift their phoneme boundaries in the direction suggested by the disambiguating context (Bertelson, Vroomen, & de Gelder, 2003; Norris, McQueen, & Cutler, 2003). Similar perceptual learning mechanisms also appear to guide adjustments to foreign-accented speech. Using a paradigm similar to that used by Norris, McQueen, and Cutler (2003), Eisner, Melinger, and Weber (2013) showed that native English listeners can adjust to word-final devoicing of obstruents in Dutch-accented English. When exposed to auditory English words that did not form words if the final consonant was interpreted as voiceless (e.g., "overload") during exposure, listeners retuned their

phonetic categories to adjust to the speaker's accent. At test, these listeners interpreted words with final devoicing correctly, even though these words now formed minimal pairs with words ending in /d/ (e.g., "seed" produced as [sit] was interpreted as "seed" rather than "seat"). This suggests that native listeners can adjust to at least some segmental variation in foreign-accented speech, using similar mechanisms as in adjusting to native segmental variation.

Listeners not only learn about a particular speaker's accent but can also—given enough evidence to distinguish talker idiosyncrasies from accent features—adjust to a particular accent. Bradlow and Bent (2008) found that native English listeners adapted to a particular native-Chinese speaker's foreign-accented English and better understood speech produced by the same person later, but they did not show this benefit for another speaker who had the same language background. When native listeners were exposed to *multiple* speakers with the same foreign accent, however, they were able to generalize to a novel speaker with the same language background. Consequently, listeners can adjust to a particular speaker, but to adjust to an accent and generalize that knowledge across speakers, they must hear a variety of speakers with that accent. The intelligibility of the speaker, which is modulated by accent strength, thereby has an effect on adaptation: listeners were able to adjust more quickly to the speech of more intelligible speakers than to less intelligible speakers (Bradlow & Bent, 2008). Native listeners can also adjust to prosodic accent features. Reinisch and Weber (2012) examined the effects of misplaced suprasegmental lexical stress in Hungarian-accented Dutch. Native Dutch listeners who were exposed to this accent instead of canonical Dutch were able to adjust to it such that they were better able to distinguish that speaker's novel target words from stress

competitors (e.g., in English, the pronunciations of “permit”, /pəˈmɪt/ and /ˈpɪmɪt/, are stress competitors, albeit not suprasegmental ones as in Dutch; ˈ indicates the placement of primary stress). A strong body of literature thus supports the flexibility of perceptual learning about foreign-accented speech.

In the experiments conducted for this thesis, we expanded on this literature and focused on how native listeners process and adjust to phonological rule misapplications that are due to foreign accents. In Experiment 1, we tested whether native English listeners’ knowledge about a foreign speaker’s native language can help with processing a speaker’s accent feature that stems from the incorrect transfer of a phonological rule of their native language to English. The phonological rule of interest was word-final devoicing in German and its application to English. During an initial learning phase, native English listeners with minimal past experience with the German language and German accents saw printed German words while hearing them being pronounced by a native German speaker. Participants were asked to learn the spelling of these German words. During this training session, one group of participants (*experimental group*) was exposed to German words ending in the letter “g”, while another group of participants was not (*baseline group*). This orthographic “g”, with the underlying representation /g/, is devoiced and thus pronounced as more [k]-like than a non-word-final /g/ in German or /g/ in English (Port & Crawford, 1989). At a subsequent test, a cross-modal priming lexical decision task with German-accented English primes was used to determine whether exposure to this phonological rule in German had an influence on how participants perceived the intended forms of German-accented English words ending in /g/, realized as a devoiced /g/. Critically, all of these words formed minimal pairs with words ending

in /k/ (e.g., “frog”, “frock”). If participants in the experimental group learned to perceive the German speaker’s more [k]-like /g/ as “g” and applied this knowledge also to the German speaker’s accented English, then hearing a prime ending in “g” (e.g., “frog”) should facilitate responses to the same word presented as target (“frog”), compared to a condition with an unrelated prime (e.g., prime: “slip”, target: “frog”). Crucially, participants who have not been previously exposed to the German devoiced /g/ should show a smaller facilitation effect, or even an inhibition effect, if they perceive a devoiced /g/ as /k/. This result would suggest that English listeners are able to use knowledge about a phonological rule in the speaker’s native language (German) for processing that speaker’s foreign-accented speech (English). It would imply that listeners can use phonological knowledge about the speakers’ native language to help with understanding their foreign-accented speech. Alternately, if no difference between groups was found, it could be possible that either the right type or amount of learning did not occur, or that learning that did occur was limited to the language context in which it was acquired.

In Experiment 2, we examined how the failure to apply an English phonological rule in foreign-accented speech affects the segmentation of speech into words by native listeners and thus the time course of spoken word recognition and lexical competition. Unlike printed words, spoken words are not reliably separated by pauses or reliably marked by any other cue to word boundaries. Listeners thus use whatever cues are available to segment speech. One of these cues in standard American English seems to be the word-final velarization of /l/ (Nakatani & Dukes, 1977). However, German-accented speech likely only contains non-velarized instances of /l/, due to the lack of velarization

of /l/ in most dialects of German. If German-accented English /l/ is not fully velarized, then this should affect lexical segmentation by native English listeners.

Previous work has shown in two-alternative forced choice tasks that listeners of American English can distinguish ambiguous word sequences, such as “knee#lax” and “kneel#axe”, based on velarization (Nakatani & Dukes, 1977). These studies relied, however, solely on so-called *offline* tasks. Offline tasks provide a measure of the end product of the recognition process, often even giving listeners unlimited post-perceptual processing time to arrive at these end products. It is therefore unclear from the previous literature whether velarization is a cue used *during* recognition and/or at a later, post-perceptual or decision-related stage. In contrast, *online* tasks tap into the recognition process as it unfolds and can thus inform whether a cue is indeed used during recognition. Reinisch and colleagues (Reinisch, Jesse, & McQueen, 2011) have contrasted the use of speaking rate to interpret duration as a lexical segmentation cue in *offline* versus *online* tasks. Dutch listeners in this study applied information about the speaking rate of the preceding context to interpret a durational cue to segmentation. Listeners relied more on the rate of the immediately preceding context than on the rate of distal context during word recognition, but during post-perceptual processing, the rate of distal context became more important than the proximal rate. These results show the importance of considering the nature of the tasks in interpreting the use of cues. As a first step towards examining whether the failure to velarize in German-accented English results in a segmentation problem for native English listeners, we first tested whether and when velarization in native English speech is used as a cue for word recognition. To this end, we tested its use in both an online eye-tracking task using the visual world paradigm and in an offline

categorization task. Following Reinisch et al. (2011), we used the printed version of the visual world paradigm, in which listeners hear an ambiguous word sequence (e.g. “knee#lax”), while seeing the second word as a target (“lax”), a competitor that is phonologically related to the alternative segmentation (“act”), and two phonologically and semantically unrelated distractors (e.g., “quick”, “pooch”). While listening to these sequences, the probability of listeners spontaneously fixating these printed words relates to the degree to which listeners momentarily considered a word to be a viable candidate (Allopenna, Magnuson, & Tanenhaus, 1998; Cooper, 1974; Huettig & McQueen, 2007; McQueen & Viebahn, 2007; Reinisch, Jesse, & McQueen, 2010; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). If English listeners use the presence and absence of velarization as a cue during word recognition, then they should be able to look more at the target than at the competitor before the target becomes segmentally unique.

In Experiment 2, we will thus establish whether velarization is used as a cue during online word recognition, which no study has directly tested to date. The above experiments will set the basis to conduct a similar experiment with German-accented speech stimuli in place of native English speech. This experiment, discussed in more detail in a later chapter, will test how the language system deals with a missing (or not fully realized) but expected cue for lexical segmentation and whether it can recover and learn to recognize the target before the point of disambiguation. Overall, the results from this set of experiments will thus demonstrate whether listeners use velarization of /l/ as a segmentation cue, and more broadly, whether listeners improve their segmentation ability with repeated exposure to a foreign accent that fails to apply the phonological rule underlying the cue.

In sum, the experiments proposed for this thesis inform us about the processes underlying how listeners understand and learn about phonological rule misapplications in foreign-accented speech.

CHAPTER II

EXPERIMENT 1

Phonetic and phonological features are perhaps some of the most common and readily apparent components of foreign speech. These divergences from native speech can be attributable to differences in the phoneme inventory of the L2 and L1 language. Substitutions of native phonemes for non-native ones are common in second language production. For example, German's inventory does not contain the phonemes /θ/ and /ð/, so native German speakers often produce /s/ or /z/ in their place, respectively (Howell & Dworzynski, 2001). Similarly, Dutch speakers most commonly replace /θ/ with /t/ (Wester, Gilber, & Lowie, 2007). Accordingly, Germans perceive /s/ rather than /t/ as acoustically more similar to /θ/, while Dutch perceive /t/ rather than /s/ as more similar to /θ/ (Hanulíková & Weber, 2012). Non-native speakers of English thus replace a phoneme of their L2 with the “closest” existing native phoneme (Flege, Schirru, & MacKay, 2003). Even when phonemes are not entirely substituted with native ones in producing an L2, they often vary from how native speakers would produce them. As mentioned above, L2 speakers may produce sounds in their L2 with subphonemic features of their L1 (Busà, 1992, as cited in Munro, Flege, & MacKay, 1996). For example, native speakers of Spanish produce the voiceless plosives /p, t, k/ with a shorter, more Spanish-like voice onset time than native English speakers (Flege & Eefting, 1987).

Another reason that phonemes in foreign-accented speech can differ from native speech is that non-native speakers also often apply phonological rules from their native language to an L2 or fail to apply phonological rules in the L2 because they do not exist in their native language. A phonological rule of German and Dutch is word-final

obstruent devoicing, in which voiced obstruents become voiceless in word-final position (e.g., /hɒnt/, “dog”, but /hɒndə/, “dogs”). Previous work has shown that German minimal pairs of words with devoiced, word-final /d/ or word-final /t/ have subtle differences, such as in preceding vowel duration and burst duration, but these differences are not always consistently produced even within the same speaker’s utterances (Port & Crawford, 1989). Because of these subtle differences German listeners can distinguish between word pairs such as *Rad*, “bicycle” and *Rat*, “advice” (which have a devoiced /d/ and a /t/ respectively at their end) above chance, but not perfectly. It would thus follow that similar phonetic differences apply to word-final /g/ and /k/ in German. Native German and Dutch speakers apply this phonological rule also when speaking English as a second language, producing weaker acoustic cues for voicing word-final obstruents than native English speakers (Simon, 2008; Smith, Hayes-Harb, Bruss, & Harker, 2009). In this case, a phonological rule is being applied cross-linguistically in a foreign accent. In other instances, a phonological rule of the L2 may not be applied because the rule does not exist in the L1. For example, when /t, d/ are intervocalic in English, these phonemes are realized as the flap [ɾ]. However, Arabic does not have this flapping rule even though its inventory contains [ɾ], and as a result, native Arabic speakers produce /t, d/ intervocalically instead of the flap (Flege & Port, 1981).

Differences from native speech like these can make foreign-accented speech more difficult to process, but listeners can adjust with exposure. Eisner, Melinger, and Weber (2013) examined how native English listeners learned about the application of the word-final obstruent devoicing rule by Dutch speakers to English. In an auditory lexical decision task, listeners were exposed to Dutch-accented English that either contained

tokens with devoiced final obstruents such as “overload” or not. Critically for inducing lexically-guided perceptual learning, these tokens did not form a real English word if the /d/ was interpreted as /t/ (e.g., “overload” is not a word). Following that, a cross-modal priming lexical decision task, again using Dutch-accented English, demonstrated that only participants who were exposed to the devoicing showed facilitatory identity priming for pairs such as auditory [sit] followed by visual “seed”, in which the prime was a word that formed a /d, t/ minimal pair (e.g., “seed” and “seat”). Participants who did not receive exposure to this feature of Dutch-accented English did not show significant facilitation. Listeners were thus able to adjust to this accent feature. This study examined learning from exposure to the foreign-accented speech. But, can listeners learn about the rule in the speaker’s L1 and use that knowledge to better understand the L2? Or is this knowledge limited to the language context in which the exposure occurs?

The goal of the present study was to examine whether native speakers of English can learn about a phonological rule in German implicitly through training on German spelling with exposure to spoken German words, and apply it to German-accented speech. As discussed above, word-final devoicing of obstruents is a phonological rule in German, and German speakers tend to apply it to their speech in English, thus producing tokens that sound less voiced than when produced by English speakers. For example, German-accented “bead” would be closer to [bit], and “frog” would be closer to [frɔk], than their native voiced coda counterparts. The latter example, final devoicing of /g/, is the critical case of this rule used in the following experiment. If native English listeners can learn about this phonological rule through exposure to German—seeing that in German, an orthographic “g” at the end of a word is pronounced more like [k]—they may

be able to use that knowledge to adjust to and more easily understand the incorrect application of this rule in German-accented English. We conducted a three-phase experiment, consisting of a German spelling training phase, an English test phase, and a German spelling post-test. During training, native English participants with no prior knowledge of German and little exposure to German-accented English were instructed to learn the spelling of the German words they heard. Orthographic representations of these words were provided. For some participants, the training stimulus set contained instances of a devoiced /g/ (*experimental group*), and for others it did not (*baseline group*). Exposure was followed by a multiple-choice test, in which participants had to select the correct spelling of a word. This sequence of one exposure and one multiple-choice test block was then repeated one more time. The learning phases ended with a spelling test, in which participants were asked to type the auditorily presented words using a keyboard. This test was first given with feedback, then without feedback. Only participants who passed an a priori determined learning criterion in the second multiple choice and spelling test without feedback were allowed to continue. Participants were told that they would need to remember the spelling of these words for a later spelling test, and that they would perform an intermediate task so that they could not rehearse the spelling. In reality, this intermediate task was the true test phase. A cross-modal priming lexical decision task with English stimuli was used, similar to that of the test phase of Eisner, Melinger, and Weber (2013), discussed previously, and to that of Sjerps and McQueen (2010), who showed that listeners can learn to interpret a foreign phoneme as a variation on a native sound (in this case, Dutch listeners learned to perceive /θ/ as /f/ or /s/, depending on exposure). Critical items were minimal pairs based on the voicing of the velar stop at the

end of the word (e.g., “frog”/“frock”). If participants in the experimental group learned the German phonological rule of final devoicing and could use it to adjust to its incorrect application to English, responses to the critical “g”-targets should be faster and more accurate when these targets are preceded by related auditory primes (e.g., “frog”-“frog”) than by unrelated control primes (e.g., “slip”-“frog”). In contrast, the baseline group should not show facilitation to the same extent and may even show inhibition, if the auditory prime is interpreted as a lexical competitor (e.g., if “frog” is interpreted as “frock”; Marslen-Wilson, 1990). However, if participants in the experimental group either do not learn the phonological rule, or are not able to apply it cross-linguistically, we do not expect to see differences between the two training groups. A German spelling test was given at the end of the experiment to test whether participants remembered what they had learned.

Method

Participants

Fifty-one undergraduate students (41 females, 5 left-handed, with a mean age of 20.10 years) were recruited from the University of Massachusetts Department of Psychological and Brain Sciences participant pool and received course credit for their participation. All participants were monolingual native English speakers with minimal exposure to the German language and German accents in English, as established with a survey about their language experience (see Appendix). All participants reported normal hearing and normal or corrected-to-normal vision.

Materials

Language questionnaire. A short survey was developed to assess experience with and attitudes toward German and other languages. Participants were asked to report whether they had learned a foreign language, and if so, to provide information about how and when the language had been learned and to rate their current proficiency. Participants were also asked about their exposure to the German and German-accented English (e.g., travel to German-speaking countries, exposure to German or German-accented English in their daily lives, attitudes toward the German language and culture and language learning in general). Participants with more than minimal exposure to the German language or German-accented English were disqualified from participating. Participants were excluded from participation if they reported: having learned to speak any German, having traveled to an area where a substantial part of the population speaks German (e.g., Germany, Austria, Switzerland, Belgium), watching or listening to German-language media more than “rarely”, or knowing a native German speaker who did not sound like a native English speaker. All included participants reported that to the best of their knowledge, they had never heard German-accented English in person.

Training phase stimuli. Twelve monosyllabic German words were selected as targets for the critical trials. Six of these words ended with the letter “k”, pronounced /k/. The other six words ended with the letter “g”. This orthographic “g” is realized in word-final position as a devoiced /g/, which is highly similar to [k]. In line with previous studies on perceptual learning that have shown that listeners do not retune phonetic categories for ambiguous stimuli if they also receive both endpoints of the continuum from which the ambiguous sounds were generated (Kraljic, Samuel, & Brennan, 2008),

none of the words contained the letters “k” or “g” or the phonemes /g, k/ elsewhere. For this reason, we did not expect any participants to learn the actual phonological rule, but rather that word-final “g” is produced as a [k]-like sound. An additional 12 monosyllabic German words without the letters “g” or “k” or the phonemes /g, k/ were selected as fillers. None of the filler words ended in a devoiced plosive or fricative or a different devoiced stop. This set of training phase stimuli was used for all three parts of training (i.e., for exposure, verification, and criterion check).

	<u>Related Pairs (64)</u>		<u>Unrelated Pairs (64)</u>	
	Auditory Prime	Visual Target	Auditory Prime	Visual Target
Critical (32)	snag	snag	flute	snack
	muck	muck	whale	mug
Word Filler (32)	rim	rip	E: slug B: spike	frown
	mop	mop		
Nonword Filler (64)	E: drug B: spook	E: drull B: spoot	E: smug B: fluke	brap
	flour	flouch		

Table 1. Examples of test phase primes and targets. Above, “E” denotes the experimental group, while “B” denotes the baseline group. The symbol | indicates that one option was presented to a given participant based on their group assignment.

Test phase stimuli. A total of 208 English prime-target pairs were created. Of these pairs, 96 had nonword targets and 112 had word targets. These pairs were assigned to lists, creating 128 trials per list (see Table 1). Half of all trials in each list had nonword target, and half had a word target. All prime-target pairs were monosyllabic and prime and target shared the same syllable structure within a pair (e.g., prime: “whale”, target:

“mug” (CVC); prime: “flour”, target: “flouch” (CCVC). No stimulus contained a final voiced stop other than /g/, nor did any stimulus contain /g, k/ elsewhere in the word, unless otherwise specified.

Sixteen monosyllabic English minimal word pairs that formed a word when ending in /g/ and /k/ (e.g., “snag” and “snack”) were selected as targets for critical trials. Each word in these pairs was presented as a target to each participant, for a total of 32 critical trials. One item of a pair was presented as its identity prime and target (e.g., prime: “snag”, target: “snag”; prime: “muck”, target: “muck”) and the other item in the pair was the target following a phonologically and semantically unrelated prime (e.g., prime: “flute”, target: “snack”; prime: “whale”, target: “mug”). Half of the words in each priming condition (i.e., 8 words) ended in “g” (referred to hereafter as “G words”) and half in “k” (referred to hereafter as “K words”). The assignment of words to priming condition was counterbalanced across two lists. The same word was used as unrelated prime for both members of a pair. The unrelated primes and targets in these lists were respectively matched in their average raw spoken word frequency, taken from the Corpus of Contemporary American English (Davies, 2008; List A: unrelated primes ($M = 409$, $SD = 730$) vs. targets ($M = 376$, $SD = 736$), $t(62) = 0.175$, $p = .86$; List B: unrelated primes ($M = 346$, $SD = 316$) vs. targets ($M = 376$, $SD = 130$), $t(62) = -2.18$, $p = .83$).

Thirty-two monosyllabic English words that did not contain /g, k/ (e.g., “frown”) were selected as printed targets for word filler trials. Half of these words were preceded by phonologically and semantically unrelated auditory word primes, and half were preceded by related word primes. The 16 unrelated filler primes ended in /g/ (e.g., “beg”) for the experimental group, and ended in /k/ (e.g., “beak”) for the baseline group. None of

any of these primes formed a word when the final /g/ was substituted with /k/ or vice versa. Primes ending in /g/ or /k/ were selected here so that the presence of these sounds in the primes would not be predictive of whether a related or unrelated target would follow. /g/ was not presented to participants in the baseline group to avoid adaption to devoicing during the test phase. For the related word filler trials, eight monosyllabic English word pairs were selected as primes that formed a minimal pair with their target (e.g., prime: “rim”; target: “rip”). These primes did not contain /g, k/ and did not end in a voiced obstruent. For the other eight related word filler trials, the target word was also used as a prime (e.g., prime: “mop”; target: “mop”).

For nonword filler trials, 80 phonotactically legal nonwords were created as targets. Nonwords were generated from merging two monosyllabic English words overlapping in their vowel (e.g., “chire” was formed by combining *child* and *fire*). A linguistically trained, native speaker of English checked the phonotactic legality of these nonwords. Each participant received 64 of the 80 nonwords as targets. Forty-eight of these targets were the same across baseline and experimental groups, and 16 differed across groups in order to avoid exposure to /g/ in the baseline group. Half of all nonwords were preceded by a phonologically related and half by an unrelated prime. Half of the word primes in the related condition contained a final velar stop. This velar stop was /g/ for the experimental group (e.g., prime: “drug”, target: “drull”) and /k/ for the baseline group (e.g., prime: “spook”, target: “spoot”). These words did not form a minimal pair if the final sound was replaced with the respective other velar stop for the baseline group. However, due to the limited availability of monosyllabic words ending in /g/, only half of these words did not form minimal pairs with /k/ for the experimental group (e.g., “log”).

The other half of the primes in the related condition did not contain /g/ or /k/ and were thus the same for both groups (e.g., prime: “flour”, target: “flouch”). A similar approach was used for the unrelated condition, in that half of the primes contained a velar stop (/g/ for the experimental condition (e.g., prime: “smug”, target: “brap”) and /k/ for the baseline condition (e.g., prime: “fluke”, target: “brap”)) and the other half did not (e.g., prime: “moat”, target: “chire”). The frequency of the word primes used for nonword target trials was as close as possibly matched to the frequency of word primes used for word target trials. Given the limited availability of monosyllabic words ending in /g,k/ though, primes on word trials had a significantly lower raw frequency (experimental condition: $M = 378$, $SD = 551$; baseline condition: $M = 380$, $SD = 514$) than primes on nonword trials (experimental condition: $M = 1124$, $SD = 2427$; $t(68.22) = -2.41$, $p = .02$; baseline condition: $M = 1121$, $SD = 1770$, $t(71.53) = -3.24$, $p = .002$).

Recordings and stimuli editing. All auditory stimuli were produced by a female native German speaker who had lived in Germany for the majority of her life and had learned English as a second language. She was on a short-term visit to the United States at the time of the recording. The speaker was naïve to the purpose of the experiment. German and English auditory stimuli were recorded at the end of the English carrier sentence, “The item is _____.” A token of this carrier sentence that had the same duration as the average duration of all carrier sentences was chosen and spliced before the German word stimuli. We chose to do so in order to give participants a small amount of exposure to definitively German-accented English speech from the same speaker before the test phase, so that they would be aware that the speaker was the same person in each task. Previous experiments have also indicated that retuning to voicing variability in stops is

transferred across speakers (Kraljic & Samuel, 2006), so even if participants were unaware that the speaker was the same, they would still be likely to transfer their learning. The recorded carrier sentence was removed from the recordings of the English word stimuli. We found the ratio of the average intensity of the carrier sentences to the average intensity of the words in the original recordings. This ratio was then applied to the selected carrier sentence to determine the target intensity for the final German word stimuli, to which they were all adjusted. The English word stimuli were normalized in their intensity.

Design and Procedure

We scripted our experiments in Python and Octave, with the latter using Psychophysics Toolbox extensions (Brainard, 1997). Visual stimuli were presented on a Dell SR2320L monitor. Auditory stimuli were presented at a comfortable listening level through Sennheiser HD 280 Pro headphones. This experiment consisted of three phases: a German training phase, an English test phase using a cross-modal priming paradigm, and a German spelling post-test. Participants were randomly assigned to either the experimental group (exposed to devoiced /g/ in German) or the baseline group (not exposed to devoiced /g/ in German).

Training phase. The training phase consisted of two repetitions of an exposure-verification block sequence and two criterion blocks. Participants received feedback on their performance in the first but not the second criterion block.

Each exposure block consisted of a random presentation of 24 German words. For the experimental group, the critical words were six German words ending in “g” (*Berg*, *Flug*, *Schlag*, *Trog*, *Zug*, and *Zwerg*) and six words ending in “k” (*flink*, *Leck*, *Prunk*,

welk, Werk, and Zweck). Participants in the baseline group were not exposed to any German words ending in the letter “g”, but only received the six critical “k”-words. All participants were also presented with filler words that did not contain “k” and “g” (e.g., *bloss, Duft, Narr*). The experimental group received 12 and the baseline group 18 of these filler words. Each exposure trial began with a display of a fixation cross for 250 milliseconds before a printed word was shown on the computer screen. After 1000 milliseconds, a German speaker pronounced the word shown on the screen at the end of the carrier sentence “The item is...”. The printed word remained on the screen during the audio playback and for 4000 milliseconds afterward. Then, the experiment automatically proceeded to the next trial. Participants were instructed to learn the spelling of each word such that they could spell it when hearing the word again later.

A verification block followed each exposure block. In each verification block, participants were tested in random order on the same 24 German words presented in the exposure blocks. Participants heard “The item is...” followed by a German word and were asked to identify the correct spelling of the auditory word from three choices shown on the screen. These response alternatives were the correct response, one response with a vowel substitution error (e.g., “Troag” instead of the correct “Trog”), and one with a consonant substitution (e.g., “Verk” instead of “Werk”) or with a consonant or silent “e” addition (e.g., “Dorne” instead of “Dorn”). Consonant substitutions/additions occurred half of the time at the beginning of the word (see “Verk” above) and half of the time at the end of the word, with the latter including silent “e” additions (see “Dorne” above). Additionally, only half of the critical words ending in “g” had misspelling options that substituted a “k” in place of “g” (e.g., “Berk” instead of “Berg”) in order to avoid

drawing attention to this phonological rule. Our justification was that it would be equally obvious to participants if all “g” letters had been replaced with “k”s or if none had been, because of how likely misspelling “g” as “k” would be. The assignment of the three response options to positions on the screen (left, center, right) was randomized on every trial. After the participant responded, or after 5000 milliseconds had elapsed, the correct response was displayed on the screen for 1000 milliseconds, and the experiment then automatically proceeded with the next trial. No trials were repeated within each block. Participants who achieved less than 75% total accuracy in the second verification block were excluded post-hoc from all further analyses.

The criterion blocks were presented after both exposure-verification sequences were completed. In each of the two criterion blocks, participants listened to each of the 24 previously presented German words and were asked to type in their orthographic representation using a keyboard. This part of the experiment was conducted using a simple terminal input. On a given trial, participants would first see, “Enter the spelling below then press enter:” and then the audio would play. On audio offset, the text, “SPELLING:” would appear, and participants could type in their response. There was no response deadline. After pressing enter to submit the response, the text, “Answer recorded” was displayed, and the next trial would begin. In the first criterion block, after participants submitted their response for each word, the audio would play again, and then the correct spelling of the word was presented on the screen below the participant’s response. This feedback was intended to help participants compare their responses against the correct answer. In the second criterion block, participants did not receive this feedback; the experiment automatically proceeded to the next trial after a response had

been submitted. Critical trials were the words ending in “g” for the experimental group and the words ending in “k” for the baseline group. Participants who spelled fewer than 75% of instances of the final letter (g or k) on critical trials with the correct letter in the second criterion block were excluded post-hoc from further analyses.

Test phase. Participants were tested in a cross-modal priming paradigm with foreign-accented English stimuli spoken by the same German speaker as during exposure. Participants were instructed that this part of the experiment was an unrelated intermediate task to prevent them from rehearsing the German items for a spelling test later. Each trial consisted of a foreign-accented auditory English prime played over headphones, immediately followed by a printed English target shown on the computer screen. Participants were asked to indicate by button press as quickly and as accurately as possible whether the printed target was an English word. Response labels were assigned to buttons such that participants always gave “yes” responses with their dominant hand. On a given trial, a white fixation cross on a black background was displayed for 250 milliseconds, then the cross would disappear, and the auditory stimulus would begin to play. Upon audio offset, the printed word, in white, size 60 Droid Sans Mono font, appeared in the center of the monitor with a black background. Participants then had to indicate with a button press whether the visual stimulus was a real English word. After 3000 milliseconds had elapsed or a response was recorded, the experiment continued automatically with the next trial.

Half of the trials had word and the half had nonword targets. Half of the word targets were preceded by an unrelated word prime (32 items), a quarter was preceded by a phonologically related prime (16 items), and another quarter was preceded by an identical

word prime (16 items). Half of the critical target words were preceded by a phonologically related and the other half by an unrelated prime. Half of the nonword targets were preceded by a phonologically related word prime (32 items) and half were preceded by an unrelated prime (32 items). Critical items were rotated through priming conditions. Each participant saw a total of 128 trials, presented in random order.

Spelling test phase. Last, participants performed a spelling test on the studied German words, similar to the second criterion block in the exposure phase. The only difference was that this spelling test contained all 48 items from the experimental and the baseline groups' second criterion block. Accordingly, 6 of the words in the spelling test phase were novel and had never been presented to the participant: for the baseline group, these were the 6 “g” words, and for the experimental group, these were the 6 filler words that had not been presented in their training phase. On each trial, an auditory German word was presented again following the phrase, “The item is”, and participants were asked to type in the spelling of the word. The timing of the trial events was the same as for the second verification block in the exposure phase. No feedback was provided.

Results

To ensure that participants learned sufficiently, several a priori accuracy criteria had to be met for participants to be included in the analyses. Participants had to achieve at least 75% accuracy on filler trials (e.g., for German words with no “g” or “k”) and on the critical word items in the second verification block of the exposure phase to be included in the analyses. Three participants were excluded for failing to achieve this accuracy criterion. Secondly, participants also had to achieve 75% accuracy for critical items in the second criterion check phase (no feedback given). Four participants were

excluded for not meeting this criterion. An additional four participants were excluded because they achieved less than 75% accuracy for either words or nonwords in the test phase lexical decision task, indicating that they likely misunderstood the instructions for the experiment. For the final analyses, the data of 20 participants in baseline group and 20 participants in the experimental group was analyzed.

With these criteria in place, the baseline and experiment group did not differ in their learning of the spelling of German words. The two groups showed no statistically significant difference in their overall accuracy in the second verification block (experimental group: $M = 95.62\%$, $SD = 4.96\%$; baseline group: $M = 94.37\%$; $SD = 4.12\%$; $t(38) = 0.867$, $p = .39$) and in the second criterion block (overall word accuracy, correct spelling of all letters in all word types, for experimental group: $M = 74.79\%$, $SD = 16.80\%$; baseline group: $M = 78.54\%$, $SD = 13.94\%$; $t(38) = -0.768$, $p = .45$). However, participants in the experimental group had a significantly lower mean accuracy rate for spelling the final letter of their critical words correctly (experimental: $M = 88.33\%$, $SD = 7.84\%$) than the baseline group (baseline: $M = 98.33\%$, $SD = 5.13\%$; $t(38) = -4.775$, $p < .001$), though both means were above the individual accuracy criterion of 75%. This difference is attributable to the increased difficulty of spelling the [k]-like sound with “g” for the experimental group, compared to the less difficult task of spelling /k/ with a “k” for the baseline group. Out of the 28 final letter errors in G words made by participants in the experimental group, 25 (89.29%) were substitutions of the letter “k” for “g” (the other three errors were “h”, “q”, and “r”). In comparison, participants in the baseline group only made 4 final letter errors in K words, substituting “h”, “l”, and “t”. For correct spelling of the entire word (all letters) in K words, the baseline group had a mean

accuracy of 81.67% ($SD = 17.85\%$), which was not significantly different from the experimental group's performance for those same words ($M = 78.33\%$, $SD = 24.24$; $t(38) = 0.495$, $p = 0.62$), but the baseline group's performance in spelling all letters of K words correctly showed a marginal trend of higher accuracy than that of the experimental group on overall spelling of G words ($M = 70.00\%$, $SD = 22.68\%$; $t(38) = 1.808$, $p = .08$).

Results from the final spelling test show that participants retained their knowledge. Both groups retained knowledge of the words that they had seen during exposure: the experimental group spelled the final letter in those words with "g" with 85.83% ($SD = 18.94\%$) accuracy, while the baseline spelled the final letter in K words with "k" with 98.33% ($SD = 5.13\%$) accuracy. These two means were significantly different from each other ($t(21.77) = 2.848$, $p = .009$), but did not differ from the groups' respective mean accuracy for using the correct critical final letter at the end of the exposure phase (see previous paragraph for these means; experimental: $t(19) = 0.68$, $p = .51$; baseline¹: $t(19) = 0$, $p = 1$). These results suggest that the difference between groups was attributable to the increased difficulty in spelling G words compared to K words, rather than any differences in learning. Again, the experimental group participants primarily made errors in substituting "k" for "g" at the end of G words (12 out of 17 errors), while the baseline group participants only made 2 errors (substitutions of "h" and "t"). Both groups learned similarly, and retained their knowledge similarly.

¹ Note that the accuracy results were exactly the same for the baseline group during the second criterion check phase and the spelling test phase.

		<u>Experimental Group</u>		<u>Baseline Group</u>	
		Accuracy (%)	RT (ms)	Accuracy (%)	RT (ms)
G words	Related	91.52 (9.78)	636.60 (108.97)	88.87 (14.21)	638.65 (141.43)
	Unrelated	89.76 (12.19)	682.15 (108.07)	92.23 (8.98)	697.10 (115.45)
K words	Related	92.95 (7.72)	657.95 (140.17)	88.48 (15.11)	664.10 (139.29)
	Unrelated	90.89 (11.19)	739.20 (131.64)	90.56 (11.80)	776.05 (117.27)

Table 2. Mean percentages of accurate responses and reaction times by group and priming condition. Mean reaction times above were calculated using correct word target trials only. Standard deviations are in parentheses.

The data collected for critical word trials in the test phase were analyzed in terms of accuracy and reaction time of correct responses, measured from the offset of the auditory prime. Linear mixed effects models were used to compare conditions with the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in R (version 3.1.0; R Core Team, 2014). Responses with a reaction time outside 2.5 standard deviations from the grand mean RT of correct word target responses ($M = 713.67$ ms, $SD = 282.63$ ms) were excluded from further analyses (4.16%). Table 2 shows the mean accuracy and mean reaction time for the experimental and baseline groups by priming condition. Priming effects reported below were calculated by subtracting each subject's mean for the related priming condition from that same subject's mean from the unrelated priming condition for a given type of target. This subtraction was performed separately for K words and G words.

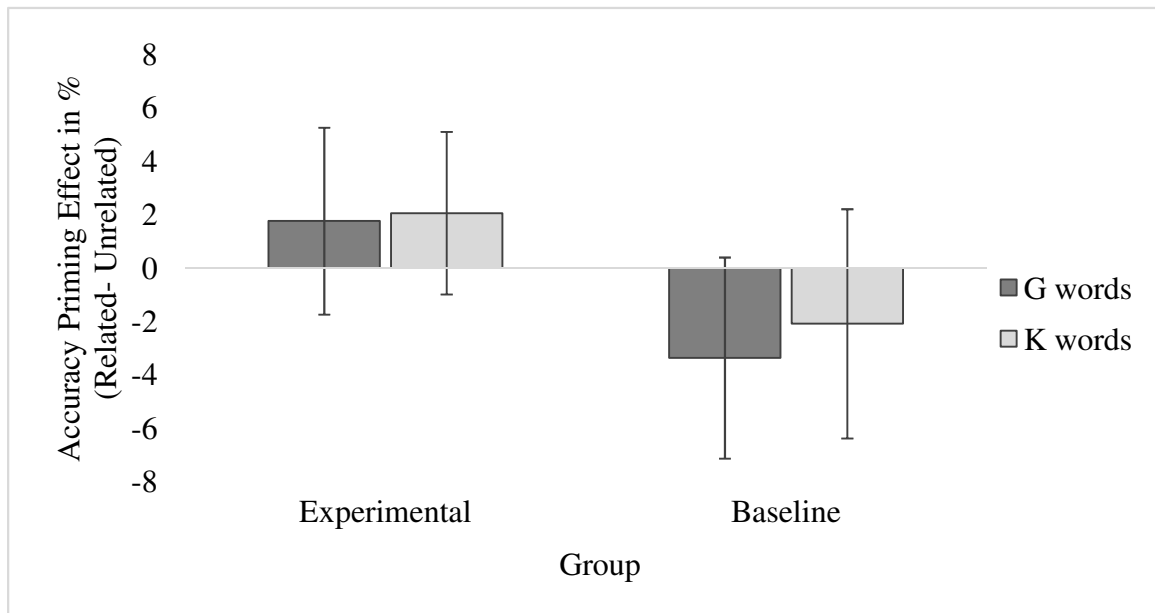


Figure 1. Accuracy priming effects in Experiment 1 for G and K words, by exposure group. Error bars represent standard deviations.

Figure 1 displays priming effect for accuracy for each critical word type by group. None of the participant groups showed a priming effect for the G or K words (experimental group: $M_g = 1.76\%$, $SD_g = 3.49\%$, $M_k = 2.05\%$, $SD_k = 3.04\%$; baseline group: $M_g = -3.36\%$, $SD_g = 3.76\%$, $M_k = -2.08\%$, $SD_k = 4.29\%$). A linear mixed-effects model examining accuracy priming, with subjects and items as random effects and word types (G or K words), group (experimental, baseline), and priming condition (related, unrelated) as contrast-coded fixed factors, revealed that only the effect of priming conditions (related vs. unrelated) was significant, showing the usual pattern of overall greater accuracy for target responses in the related than in the unrelated condition ($\beta = -1.059$, $SE = 0.355$, $p = .003$). There was no overall difference in performance between groups (baseline vs. experimental; $\beta = -0.411$, $SE = 0.400$, $p = .30$), but the interaction between group and priming condition was marginally significant ($\beta = 1.006$, $SE = 0.592$,

$p = .09$). This reflects that priming was only found for the experimental but not the baseline group, driven primarily by the experimental group's higher accuracy on related pairs. Figure 2 depicts this interaction. Word type (G or K words) had no effect and did not interact with any other factor (all $p > .05$).

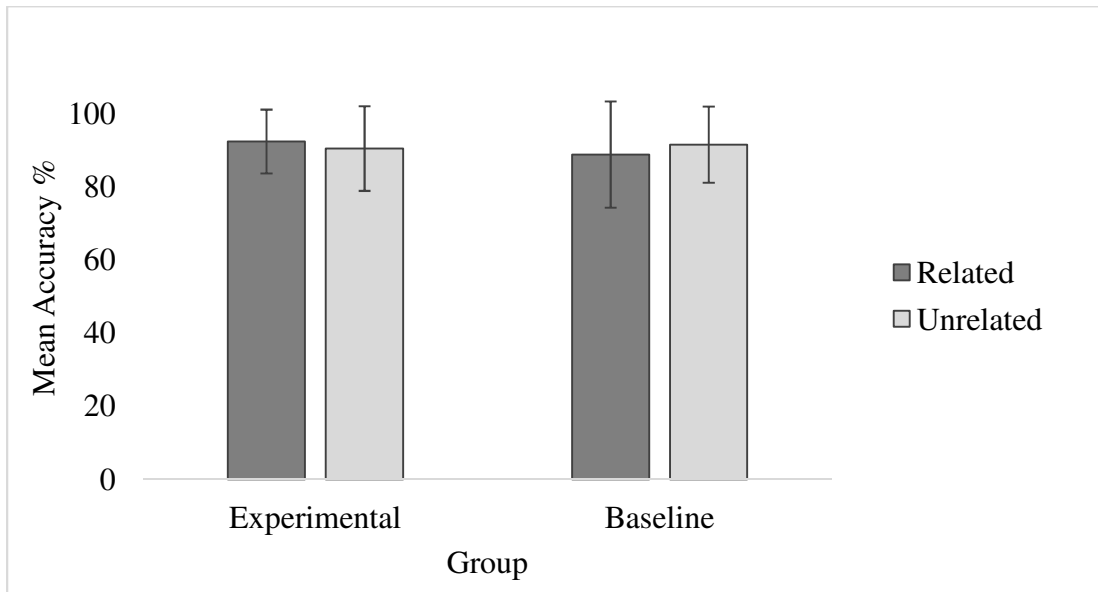


Figure 2. Mean accuracy scores in Experiment 1 for each group's related versus unrelated prime-target pairs. Error bars represent standard deviations.

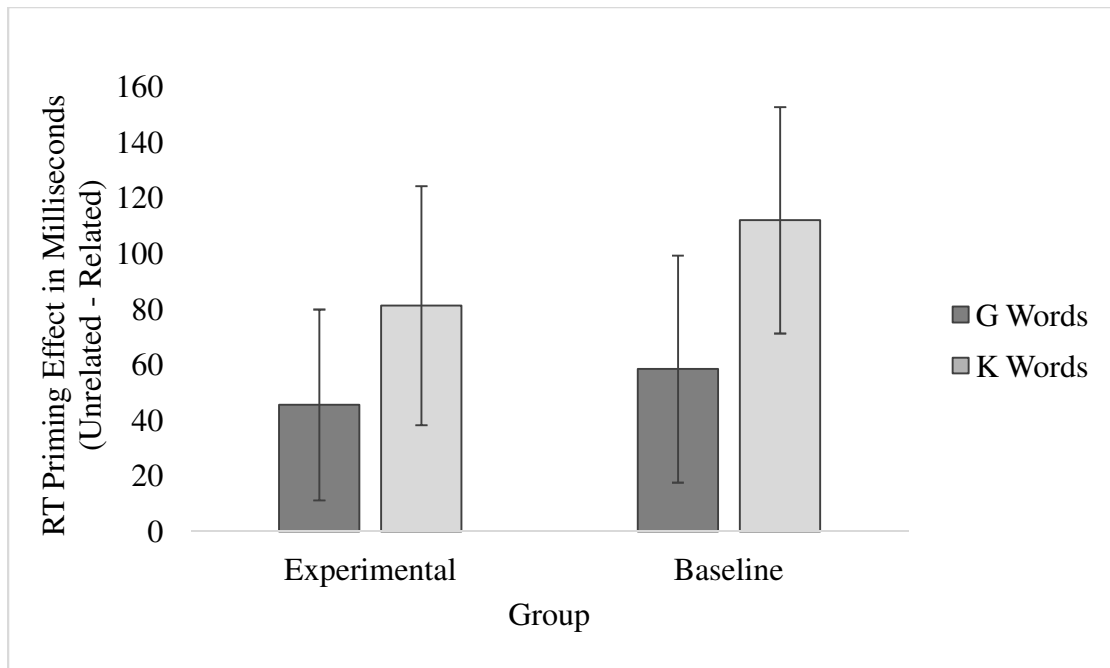


Figure 3. Reaction time priming effects in Experiment 1 for each type of critical word by exposure group. Reaction times were calculated for correct word responses only. Error bars represent standard deviations.

Figure 3 shows the mean priming effects for reaction times for each word type by group. A linear mixed-effects model using the log reaction time as the dependent variable, with subjects and items as random effects word types (G or K words), group (experimental, baseline), and priming condition (related, unrelated) as contrast-coded fixed factors, showed overall a significant facilitatory priming effect ($\beta = 0.1197$, $SE = 0.0138$, $p < .001$). Additionally, reaction times for K words were larger overall than for G words ($\beta = 0.0672$, $SE = 0.0319$, $p = 0.035$). No significant difference between groups was found, and none of the interaction effects reached the significance level, suggesting that there was no significant difference between priming effects for K and G words (all $p > .05$).

Discussion

The goal of this experiment was to examine whether listeners would apply knowledge about a foreign-accented speaker's native language to recognizing that speaker's non-native speech. English-speaking participants with no prior knowledge of German or German-accented English were taught the spelling of spoken German words. We tested whether those English speaking participants who learned about word-final obstruent devoicing in German (experimental group) during the exposure phase, specifically in the case of the phoneme /g/, were able to apply their knowledge of this phonological rule to the same speaker's German-accented English. We predicted that if so, we would see a larger facilitatory priming effect for G words for those listeners, compared to that of a second group of listeners who had not learned about the phonological rule (baseline group). If not, we did not expect to see a difference in priming for G words between these groups. We also predicted that regardless of any group differences or lack thereof, we would see facilitatory priming effects in accuracy and reaction time for K words.

The results were, however, not in line with these predictions. We found a general facilitatory priming effect in terms of accuracy and response times, demonstrating that our priming paradigm worked as intended (i.e., responses to targets were faster and more accurate if preceded by a related than an unrelated prime). However, we did not find any group differences, other than a marginally significant interaction between group and prime type for accuracy. This interaction was a marginally significant trend that the experimental group, but not the baseline group, showed facilitatory priming in terms of accuracy, but this was not modulated by item type. This could suggest that the experimental group had a somewhat better general understanding of German (i.e., not

restricted to the G items) than the baseline group, but one has to be cautious in this interpretation, since this effect is not significant and also numerically small. In sum, the results suggest that listeners in the experimental group did not apply their knowledge of word-final devoicing of /g/ in German to German-accented English with devoiced word-final /g/.

There are several plausible explanations for this outcome. First, it is possible that the experimental group did not learn the phonological rule in the first place. We used orthographic information as a method of implicitly teaching participants the rule, as it calls attention to the underlying form of the phoneme in a way that lay people would understand, i.e., “G is pronounced like a K”. The results of the criterion and the final spelling test suggest that the experimental group did indeed learn and retain the knowledge that the G words should end in the letter “g” even if they are pronounced more [k]-like than expected for a /g/ in English. However, it is possible that this orthographic knowledge did not result in the acquisition of *phonological* knowledge, even though we designed our exposure paradigms to encourage both letter-to-sound and sound-to-letter mapping. Another potential barrier to learning the phonological rule may have been the artificial format in which the words were learned, absent of semantic knowledge. The German words presented to participants were not presented with any kind of associated meaning, thus even though they were clearly words (albeit in another language), they may not have been processed in a “language-like” manner that would induce generalization of a phonological rule. Rather, the words could have been learned on a case by case basis. Another possible explanation is that the lack of *non*-devoiced, *non*-word final instances of /g/ in the German stimulus set discouraged processing of the

rule as a phonological alternation. Nonetheless, even if participants had learned that the letter “g” is *always* pronounced as more [k]-like in German, this knowledge could have theoretically been useful in processing German-accented English with this feature.

Second, our German exposure phase may not have been sufficient for the rule to be learned to the extent needed for it to be applied cross-linguistically. Due to a practical necessity, only 24 German words were presented. A previous version of the experiment with more German-language stimuli had been overwhelming to participants and prevented them from learning the words sufficiently. Out of these 24 words, only six G words were presented to the experimental group, which may not have been enough for cross-linguistic generalization. One reason why listeners may have not generalized cross-linguistically may have been because listeners had no evidence that the German speaker who produced devoicing in German would do the same in English. In comparison, in the Eisner, Melinger, and Weber (2013) study on the perceptual learning of devoicing, listeners had evidence for the accent feature in the foreign language before test, as they were trained and tested on the same speaker within the language. In our study, testing cross-linguistic perceptual learning, listeners may have assumed that devoicing is specific to the situation in which they became aware of the rule: that is, to speaking German. An option to remedy this would be to first introduce both groups to the speaker’s word-final devoicing in English before taking part in the study.

An alternative option would be to try to convince listeners to treat the accent feature as an idiosyncrasy of the speaker. In a follow-up experiment, we are currently testing this idea. The set-up of the study is similar to the one here, except that now participants hear *three* speakers during the German training phase. One of these speakers

(the same one as was used here) is the target speaker, who will then also be heard during the English test phase. This speaker devoices in both German and English. The other two speakers, however, are only heard during the training phase. In one condition of the experiments, these two speakers also devoice in German, thus conveying the idea that devoicing is an accent feature. In another condition of the experiment, these two speakers produce genuine /g/ sounds at the end of words in German (i.e., unknown to the listeners, they actually “mispronounce” the G words). This should suggest to the listener that devoicing is an idiosyncratic feature of the target speaker. We predict transfer when devoicing is interpreted as an idiosyncratic feature rather than a language feature. The addition of multiple speakers in the exposure phase could, however, also help to make the generalization more robust. While Bradlow and Bent (2008) found that participants were able to better understand an accented speaker after hearing them speak (in the same language) before, their results indicated that generalization of knowledge about an accent across speakers would only occur with exposure to multiple foreign-accented speakers. Using multiple speakers in the exposure phase could theoretically make learning of the phonological features more robust, and then perhaps more likely to foster their transfer cross-linguistically.

Third, participants may have learned the phonological rule effectively in German, but they were unable to apply it to German-accented English. This interpretation is in line with that of Levi, Winters, and Pisoni (2011), who trained monolingual English participants to identify voices of German L1, English L2 bilingual speakers, when speaking in either German or English. Participants were then asked to perform a word recognition task with the same or new speakers in English. Listeners only showed a

same-speaker intelligibility advantage at test, when hearing the speaker in the same language as during training (English), but not if they were exposed to that speaker using a different language that was unknown to the listeners (German). While our task was different in nature, both tasks require the processing of fine-phonetic detail. It is possible that the same principles underlie both results: that phonological learning is dependent on the language context in which it is presented. Our results most strongly support this possibility, but further investigation into the effects of different types of exposure is needed to draw conclusions about the transfer of phonological rule knowledge.

CHAPTER III

EXPERIMENT 2

Within a given language, phonemes vary systematically in their phonetic realization. In some cases, these allophonic alternations depend on the position of a segment in the syllable structure of a word. Listeners are sensitive to this allophonic variation when segmenting speech into words (Nakatani & Dukes, 1977). One such example of alternation in standard American English is the velarization of the alveolar lateral approximant /l/ in a syllable's coda position (Lehiste, 1964). This velarized lateral approximant [ɫ] is referred to as "dark L", while its non-velarized counterpart [l] is referred to as "light L". These allophones vary in the difference between the formants F_1 and F_2 : [ɫ] has a smaller difference between those formants than [l] (Giles & Moll, 1975). The size of this difference varies with degree of velarization. Sproat and Fujimura (1993) found that the most velarized /l/s occur in word final syllabic coda positions, especially when followed by major intonational boundaries. Given that [ɫ] is an indicator of the endings of words, it is likely that native English listeners use it as a segmentation cue.

Indeed, English listeners use the velarization of /l/ in word segmentation. Nakatani and Dukes (1977) examined cues for word segmentation with stimuli consisting of ambiguous phrases that could be interpreted in two ways depending on where the listener placed the word boundary; for example, one phrase could be interpreted as either "we loan" and "we'll own". Ambiguous phrases were created by splicing juncture phonemes from one phrase into the other. For example, the light L from "we loan" was

be spliced into word-final position to create “we’ll own” and the dark L from “we’ll own” was spliced into word-initial position to create “we loan”. Participants were asked to categorize these items as “we’ll own”, “we loan”, “we own”, or “we’ll loan”. Listeners perceived the second word as “loan” in approximately 29% of trials when hearing the cross-spliced, unvelarized [l] in “we’ll own”, and perceived the first word as “we’ll” in about 38% of trials when hearing the cross-spliced, velarized [ɫ] in “we loan”. These results demonstrate missegmentation of words due to “inappropriate” velarization in onset and the lack of velarization in coda position, suggesting that English listeners are sensitive to velarization of /l/ as a word segmentation cue.

The aforementioned experiment used speech with spliced phonemes to test participants’ perception of word boundaries. However, whether velarization is a cue to word segmentation has not been tested using natural speech. While cross-spliced speech allows for greater experimental control over stimuli, cross-spliced stimuli are less ecologically valid than natural speech. Cross-spliced phonemes may be recognized as artificial and strange due to subtle mismatches in coarticulation, which may cause participants to process these stimuli, especially their cues, differently—especially so when there is more time for post-perceptual processing, such as in an offline task. This could result in less pronounced disruption of speech segmentation. Therefore, natural speech stimuli may show effects in speech segmentation that may have been previously minimized. The second experiment in this thesis examined the use of velarization as a cue to segmentation using natural word sequences. An offline categorization task was used to investigate perception of the second word in a sequence of two words with two segmentation possibilities, for example, “teal#egg” and “tea#leg”. In addition, we used

eye tracking to track how velarization in native English speech affects the time course of word segmentation and thus lexical access and competition. Online measures of word recognition, such as eye tracking, can provide insight into when velarization affects speech segmentation and whether it is used as an immediate cue to segmentation. This experiment used the visual world paradigm, in which eye fixations on referents on a screen were recorded during speech perception. In this paradigm, the proportion of fixations to visual referents reflects the underlying lexical activation of the words that the referents represent (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Eye fixations to each item onscreen were measured with an eye-tracking apparatus. This paradigm is most often used with pictures, but the usage of printed words instead has also been demonstrated to be effective for measuring word recognition and phonological lexical competition (McQueen & Viebahn, 2007) and further, for examining the time course of word segmentation (Reinisch, Jesse, & McQueen, 2011). In the present version, participants would hear an English two-word sequence (e.g., “teal#egg”) with word-initial or word-final /l/ spoken by a native English speaker while seeing four printed words on the screen. One of the words shown on the screen was the second word in the sequence, that is, the target (“egg”), another one was the phonological competitor to the alternative segmentation (“left” for the segmentation “tea#leg”), and the other two words were phonologically and semantically unrelated distractor items. We predicted that if participants use velarization of /l/ as a segmentation cue, they would fixate already more on the target word than on the phonological competitor before the auditory input provides information to segmentally disambiguate them. In combination with natural speech

stimuli, this online measure provides new insights into how listeners use allophonic cues to segment speech.

A lack or smaller degree of velarization of word-final /l/ could be an accent feature found in non-native English. If L2 English speakers' native language does not contain an allophonic contrast, their English production is also unlikely to contain it (Flege, Schirru, & MacKay, 2003). For example, in German [l] is not velarized, particularly in southern dialects of Bavaria and Swabia (Recasens, Fontdevila, & Dolors Pallarès, 1995). We thus postulate that native German speakers who have acquired English as a second language may produce a light or less velarized L in word-final position instead of a dark L when speaking English. We therefore propose a follow-up production experiment in which we first test this hypothesis with an acoustic analysis of a native German speaker's foreign-accented English. We predict a significantly larger difference between the first and second formant (more similar to English [l] than [ɫ]) of the native German speaker's word-final /l/ in English compared to that of a native standard American English speaker. If our hypothesis about the lack of velarization in German-accented English holds, then we will conduct the same eye-tracking experiment as proposed above with German-accented English. We will then examine whether monolingual native English listeners missegment continuous speech by perceiving a word boundary before a light L in German-accented English as compared to standard American English. This effect will be shown as a modulation of the size of differences in fixation proportions to the target word versus a phonological competitor for its alternative segmentation candidate across the two speech conditions. We expect that participants hearing German-accented English will initially show a larger amount of competition (i.e.,

a smaller difference between proportions of fixations to targets and competitors) compared to participants hearing native English. We will also test whether English listeners can adjust to this accent feature. To this end, we will investigate how listeners use this ambiguous speech signal at first exposure with regards to word segmentation, as well as any disruption in lexical access and competition, and changes in processing with repeated exposure. We hypothesize that if learning about this missing segmentation cue is possible, over blocks with repeated exposure, participants in the German-accented English condition will learn to correctly segment after a word-final light L in the foreign-accented speech. This will be demonstrated by a faster rise in the difference in fixation proportions to the target compared to the competitor. However, another possibility is that the native English phonological rule will override this learning, in which case we predict that no difference in segmentation ability will occur over time for this group.

Method

Participants

Categorization. Twenty-four undergraduate students were recruited from the University of Massachusetts Psychological and Brain Sciences Department's participant pool and received course credit as compensation. All of them were monolingual native English speakers and reported normal hearing and normal or corrected-to-normal vision. All of them reported no language or attention deficits.

Eye tracking. An additional twenty-four undergraduate students from the same population as for the two-alternative forced-choice experiment were recruited as participants.

Materials

Categorization. Twenty-four critical English word pair sequences with the juncture phoneme /l/ and three sets of 24 (i.e., 72 total) filler sequences with the juncture phonemes /s/, /t/, or /k/ were selected. Each sequence was presented at the end of the carrier sentence, “On a new sheet of paper, he writes...”. Each sequence was ambiguous such that the sequence could either be segmented before or after the juncture phoneme (if velarization was not used as a cue). For example, the sequence /til#g/ could be segmented as “tea#leg” or as “teal#egg”. That is, each word in these sequences formed a word with and without the juncture phoneme. Given that each critical word pair sequence could be realized in two different ways, a total of 48 critical sequences were created. For critical word sequences, each segmentation possibility of the first word in each sequence was used twice with each critical pair of second word segmentation alternates (e.g., “kneel ash”, “knee lash”, “knee link”, “kneel ink”, “kneel eagle”, “knee legal”, “knee lax”, “kneel axe”). Both segmentation possibilities of the first word in the 72 filler sequences were used twice with each of the 72 unique filler targets (e.g., “moss creep”, “mosque able”, “mosque luck”, “moss cower”). The second word in each pair was considered the target.

The juncture phoneme was not part of an onset or coda cluster in the critical words. The juncture phoneme /l/ was always preceded by one of the following vowels: /i/, /e□/, or /a□/. The vowels following the juncture phoneme /l/ were variable. The words in each sequence ranged from 1 to 2 syllables in length. Each of the /l/ and /s/ sets,

there were nine 1 syllable second words and three 2 syllable second words of each type (word-initial juncture and word-final juncture). For each of the /k/ and /t/ sets, there were three 1 syllable second words and nine 2 syllable second words of each type. Thus, there were 24 1 syllable word-initial juncture second words, 24 2 syllable word-initial juncture second words, 24 1 syllable word-final juncture second words, and 24 syllable word-final juncture second words in a given list. The means of the possible second words in the critical pairs were matched in each stimulus list in their raw spoken word frequency from the COCA database ($M_1 = 2791$, $SD_1 = 7015$; $M_2 = 3509$, $SD_2 = 9939$; $t(23) = .23$, $p = .78$) (Davies, 2008).

Eye tracking. The same stimuli as in the two-alternative forced-choice experiment were used as auditory materials for the eye-tracking experiment. The second word in a pair was considered the target word (e.g., in the sequence “kneel eagle”, “eagle” was the target word). For each target word, four printed words were shown on the screen, consisting of the target, a phonological onset competitor for the alternate segmentation possibility, and two unrelated distractor words. All words on a display were either one syllable, two syllables, or half one and half two syllables, and matched on raw spoken word frequency within lists. Two-way 2x3 analyses of variance on the log of raw frequency confirmed that there were no significant main effects of set type (critical items, filler items like critical items, and filler items with related distractors; List 1: $F(2,372) = 0.24$, $p = .79$, List 2: $F(2,364) = .20$, $p = .82$), or displayed word type (target, competitor, distractor 1, distractor 2; List 1: $F(3,372) = 0.17$, $p = .92$, List 2: $F(3,364) = 0.33$, $p = .81$), although two sets with frequency outliers were excluded from list 2. There was also no significant interaction (List 1: $F(6,372) = 1.47$, $p = .19$, List 2: $F(6,364) = 1.20$, $p =$

.30). No items were repeated across trials. The phonological competitor overlapped in the first two segments (after the juncture phoneme, if it is part of the target word) with the target and was semantically unrelated to the target (e.g., target: “ash”, competitor: “lapse”; target: “lash”, competitor: “agile”). The distractors were phonologically and semantically unrelated to all other words in a display (e.g., target: “ash”, distractor: “fork”) and to both segmentation alternates for each word in the pair.

The same filler word sequences as in the two-alternative forced-choice experiment were presented here. These consisted of three sets of 24 fillers each, using the juncture phonemes /k, s, t/. The first set of 24 fillers, which used the juncture phoneme /k/, had the same distribution of numbers of syllables within a display set as the critical items. On these trials, just as on the critical trials, a target and a phonological onset competitor of the alternate segmentation were shown together with two unrelated distractors (e.g., auditory sequence: “croak raft”, target: “raft”, competitor: “crab”, distractors: “men”, “sip”). The remaining other 48 filler sequences with the juncture phonemes /t, s/ were included to make it unpredictable as to whether or not the target was one of the two phonologically overlapping words shown on the screen. On these trials, displays showed the target, an unrelated distractor, and two distractors that were phonologically related to each other but not to the other words shown on the screen or the auditory word pair for that trial. For example, for the auditory sequence “scarce light”, the display showed the target “light”, an unrelated distractor “banker”, and the two distractors “forge” and “organ”, that were phonologically related to each other in the same way as the targets and competitors were in the other two stimuli sets.

Recordings and stimuli editing. One native, monolingual female speaker of standard American English, naïve to the purpose of the experiment, was recorded². Target and filler word sequences were recorded at the end of the English carrier sentence, "On a new sheet of paper, he writes ____." A token of this carrier sentence with the average duration of all carrier sentences was chosen and spliced before each auditory word sequence. The intensity of the word stimuli was adjusted to the average intensity ratio of carrier sentence to word. The same auditory stimuli were used in the two-alternative forced-choice task and the eye-tracking task.

Measurement	Word-final /l/	Word-initial /l/	<i>t</i> (23)	<i>p</i>	<i>Cohen's d</i>
F ₂ -F ₁ (Hz)	729.66	882.29	4.203	< .001	0.858
Duration of /l/ (ms)	63.46	66.79	1.282	.213	-
Duration of preceding vowel (ms)	158.92	158.54	-0.087	.534	-

Table 3. Mean values of acoustic measurements and statistical tests for critical item pairs.

Acoustic analyses. An acoustic analysis of the critical tokens was conducted to confirm the larger degree of velarization of /l/ in word-final compared to word-initial position and examine the role of other potential segmentation cues. Table 3 contains the mean and statistical analyses of each measure. Previous work has shown that segment

² In addition, a female native speaker of German who learned English as a second language in school was also recorded. The productions of this speaker will be analyzed to test the feasibility of conducting future research on German-accented English.

duration can be position-sensitive, as in the case of word-final lengthening (Oller, 1973). The duration of the vowel preceding /l/ and /l/ duration were selected in order to account for the potential use of these features as cues for word segmentation. A larger difference between F_1 and F_2 is associated with [l], while a smaller difference is associated with the velarized [ɫ]. A paired samples t-test showed that this F_2 - F_1 difference was significantly greater in word-initial /l/s ($M = 882.29$, $SD = 127.64$) than word-final /l/s ($M = 729.66$, $SD = 133.20$), thus confirming that the word-final /l/s were more velarized than the word-initial /l/s ($t(23) = 4.203$, $p < .001$). No significant difference was found between items with word-initial /l/s and word-final /l/s for the duration of /l/ ($t(23) = 1.282$, $p = .213$). Additionally, we measured the duration of the vowel preceding each /l/, and no significant difference was found between conditions ($t(23) = -0.087$, $p = .534$).

Design and Procedure

Categorization. Participants were seated in front of a computer screen in a sound-attenuated IAC booth and listened to the auditory stimuli at a comfortable hearing level through a pair of Sennheiser HD 280 Pro headphones. On each trial, a fixation cross appeared for 250 milliseconds in the center of the screen, then one of the auditory word pairs, preceded by, “On a new sheet of paper he writes _____”, was played. Printed representations of the two segmentation possibilities of the target word (e.g., for auditory “knee lax”, “lax” and “axe”) were then presented on the screen, and participants were asked to indicate with a button press which of the two words they perceived as the second word in the pair. The experiment continued with the next trial after the participant responded or 5000 milliseconds elapsed. Trial order was fully randomized.

Half of all word sequences were presented with the juncture phoneme as the initial phoneme of the second word in the sequence (e.g., “leg”), and half were presented with the juncture phoneme as the last phoneme of the first word in the sequence (e.g., “egg”). The assignment of word sequence to these conditions was counterbalanced across participants. Additionally, targets containing the juncture phoneme and targets lacking the juncture phoneme were equally likely to be monosyllabic or disyllabic. Response alternatives beginning with the juncture phoneme were always presented on the left side, and response alternatives not beginning with the juncture phoneme were presented on the right.

Eye tracking. Participants were seated 60 cm in front of a Dell SR2320L monitor in a sound-attenuated IAC testing booth and listened through all auditory stimuli at a comfortable listening level through a pair of Sennheiser HD 280 Pro headphones. Participants’ eye movements were recorded at a sampling rate of 1,000 Hz with a desktop-mounted SR Research Eyelink 2000 system. The experiment was controlled by the Experiment Builder software (SR Research). Drift corrections were done on every 8th trial.

On each trial, a fixation cross first appeared for 250 milliseconds, followed by a display with four printed words onscreen in white, in a size 20 Lucida Console monospaced font, each centered in one of the quadrants of the screen on a black background. Across trials, targets and competitors occurred equally often in each position in each condition. Participants then heard the phrase, “On a new sheet of paper, he writes” continued with a pair of words such as “teal#egg”. The onset of the auditory stimuli was timed such that the onset of the juncture phoneme occurred 2,000 ms after the

onset of the display. Participants were asked to click with a computer mouse on the final word that they heard in the sentence from among the items presented onscreen.

Three blocks of 32 trials were presented, each containing 4 trials with critical targets with /l/ in onset position (e.g., “leg”), 4 critical targets with no /l/ in onset position (e.g., “egg”), and 24 filler trials containing a target with a different juncture phoneme. The order of these 3 blocks was counterbalanced across subjects, following a Latin square design. Trial order within each block was randomized. The assignment of word sequences to these conditions was counterbalanced across participants. Targets that began with the juncture phoneme and targets that did not were equally often monosyllabic and disyllabic. Half of the displays contained only monosyllabic or disyllabic words (i.e., all words were the same number of syllables), and the other half contained two monosyllabic words and two disyllabic words. Segmentation minimal pairs (e.g., “lynch” and “inch”) never occurred onscreen together.

Results

Categorization

Behavioral data from the categorization task revealed a significant bias toward assuming the juncture phoneme /l/ was word-initial. A paired samples t-test comparing accuracy for these two conditions in critical items showed that this difference was statistically significant ($t(23) = -11.15, p < .001$). The mean accuracy for critical target words not beginning with /l/ was 33.68% ($SD = 18.30\%$), while the mean accuracy for critical target words beginning with /l/ was 90.63% ($SD = 11.61\%$). A one-sample t-test comparing the mean accuracy of critical targets starting with /l/ to chance performance

(50%) was significant, indicating that participants performed better than chance on those items ($t(24) = 17.15, p < .001$). However, participants also performed significantly less accurately than chance with critical targets not starting with /l/ ($t(24) = -4.37, p < .001$). For critical targets beginning with /l/ (less velarized), participants indicated that the word began with /l/ on 90.94%³ of trials. However, for critical targets not beginning with /l/ (more velarized), participants indicated that the word began with /l/ on 66.32% of trials. The difference between these two percentages was significant, indicating that although there is a possible bias, there was still an influence of the degree of velarization on segmentation ($t(23) = -6.78, p < .001$).

Eye Tracking

Practice trial data were not analyzed. A total of 49 trials of the main part of the experiment were excluded from all analyses. Nineteen of these trials (0.82% of all trials) were excluded because participants clicked with the computer mouse outside of the defined response areas for each word (see below for criterion). Another 29 trials (1.26%) were excluded due to off-screen fixations. Finally, one trial (0.04%) was excluded because the trial timed out before a response had been given.

Mouse click responses were scored as either being on the target, competitor, or either distractor if they were within a Euclidean distance of 120 to the onscreen coordinates of each printed word. Overall, as expected, participants' click responses were highly accurate. For critical trials, participants selected the correct target in 95.83% of

³ This percentage differs from the percentage of correct word-initial trials because there was one trial that timed out before any response was given. This trial was considered incorrect for accuracy purposes, but was not considered to have a response, and was thus excluded from the analysis of response proportions.

trials. The remaining 1.74% of the responses were to the competitor. Filler trial performance was similarly accurate, with participants selecting the target on 99.34% of trials.

Eye movement data were grouped into 750 time bins of 4 milliseconds. We defined a fixation as on one of the four words onscreen if it was within a Euclidean distance of 194 from the coordinates of each printed word. The critical time window for analysis was defined as from 200 milliseconds after target onset to 200 milliseconds after the mean of the point of disambiguation for each condition (word-initial, $M = 305$, versus word-final /l/, $M = 231$). A time lag of 200 ms was selected in order to account for the estimated average 200 milliseconds required to program and initiate a saccade after the relevant auditory information has been presented (Hallett, 1986; Matin, Shao, & Boff, 1993).

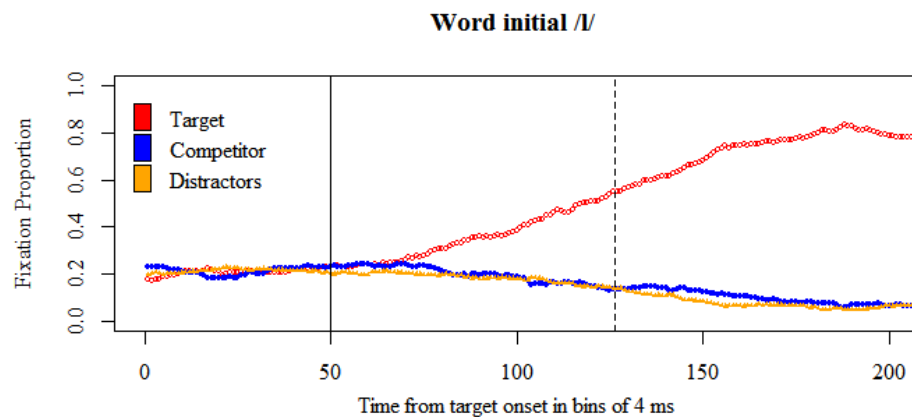


Figure 4. Fixation proportions to targets, competitors, and distractors in the word-initial /l/ condition. The solid line marks 200 milliseconds after juncture phoneme onset, while the dashed line indicates the mean point of disambiguation for word-initial /l/ trials, shifted by 200 milliseconds.

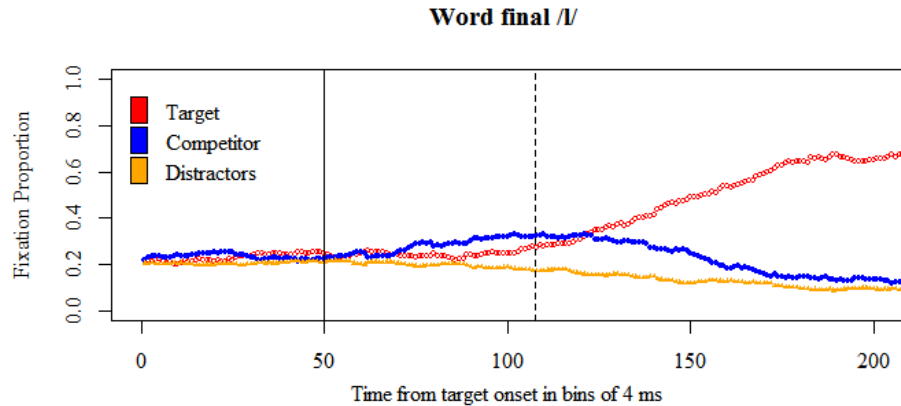


Figure 5. Fixation proportions to targets, competitors, and distractors in the word-final /l/ condition. The solid line marks 200 milliseconds after juncture phoneme onset, while the dashed line indicates the mean point of disambiguation for word-final /l/ trials, shifted by 200 milliseconds.

Eye movements for critical trials with /l/ as juncture phoneme were analyzed using linear mixed effect models with the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in R (version 3.1.0; R Core Team, 2014) examining the empirical logit of the proportion of fixations (Barr, 2008). Participant and item were used as random factors, while condition was used as a numerical fixed factor, with word-initial /l/ trials coded as 0.5 and word-final /l/ trials coded as -0.5. The dependent variable was the competition effect was the difference in the logit-transformed proportions of fixations on the targets and the competitors during the critical time window. Analyses were done separately by juncture phoneme position condition. Figure 4 shows the proportion of fixations to the target, competitor, and distractors over time for word-initial /l/ trials. For word-initial /l/ trials, participants showed a clear bias toward fixation on the target versus the competitor, or significantly reduced competition, $\beta = -1.455$, $SE = 0.406$, $p = 0.002$. Figure 5 shows the proportion of fixations to the target, competitor, and distractors over time for word-final /l/ trials. However, for word-final /l/ trials, no such effect was found,

indicating that the target was not preferentially fixated upon during the critical time window for these stimuli, $\beta = -0.272$, $SE = 0.402$, $p = 0.506$. No significant difference was found between the proportion of looks toward the competitors nor distractors on critical trials in either condition, and no competitor effect was found on word-final /l/ trials (all $p > .05$).

Discussion

The goal of Experiment 2 was to establish whether native English speakers use the position-sensitive, allophonic alternation of /l/ as a word segmentation cue, and if so, when they use this cue during word recognition. The acoustic analyses of our critical stimuli replicated previous evidence suggesting that the velarization of /l/ is position-sensitive, in that word-initial /l/ had a larger difference between the first and second formants than word-final /l/, consistent with phonetic descriptions of “light” and “dark” /l/. Given that our stimuli had the intended properties, we then examined whether velarization helps word segmentation, thereby also looking at whether this cue is immediately used in segmentation and in resolving lexical competition.

In our categorization experiment, participants were heavily biased toward interpreting the /l/ in our critical stimuli as word-initial. When /l/ was not velarized, participants were very unlikely to respond that the juncture phoneme was at the end of the first word (e.g., “kneel eagle”), and strongly preferred the response in which the /l/ was at the beginning of the second word (e.g., “knee legal”). However, when /l/ was velarized (which should indicate its word-final position), participants were also significantly more likely to respond that the /l/ was the first phoneme in the second word, despite the velarization of that phoneme. In addition to this bias, there was some evidence

that velarization was used as cue for segmentation. Listeners reported /l/-initial targets more often when actually presented with /l/-initial targets than when presented with vowel-initial targets.

The bias toward categorizing /l/ as word-initial, regardless of velarization, that was found in the categorization task was not present in the eye-tracking data. Participants were able to disambiguate a word-initial /l/ before the point of disambiguation, but they maintained both possible representations before the point of disambiguation for word-final /l/, thus showing that in the early stages of word recognition, there is no bias to interpret the velarized /l/ as word-initial. We suggest that this is due to the differential ability of online (eye tracking) and offline (categorization) tasks to tap into processing at various stages of time, as also illustrated in Reinisch, Jesse, and McQueen (2011). The categorization task reflects the combined result of earlier perceptual and later post-perceptual processing at a decision stage, while eye tracking shows the time course of processing and can thus separate earlier processes in lexical access from later ones. This could suggest that velarization is a cue used early during word processing, and also that at the same time the bias seen in the categorization task emerges late in word processing, and is not present earlier on. The categorization data also support the findings of our analyses of the eye-tracking data, in that when participants encountered a velarized, dark /l/ in word-final position, they maintained both potential representations (word-initial and word-final) during lexical access, and at the decision stage, demonstrated inaccurate categorization due to a bias towards word-initial /l/. The presence of velarization in /l/ in word-final positions was therefore not used as the primary cue to denote a word boundary after this segment, while the absence of velarization in /l/ word-initial positions was

indeed used by participants in lexical access to determine to which word the phoneme belonged.

Several possibilities could account for this pattern in which listeners segmented sequences with word-initial /l/ better than sequences with word-final /l/. One such possibility is that another cue may have, however, contributed to this result: Native English speakers may have expected glottalization of word-initial vowels, given the prevalence of this feature in this position in American English (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996). For /l/-initial targets, the lack of word-initial vowel glottalization would suggest the same segmentation pattern as the lack of velarization in /l/, leading to a relatively unambiguous perception of the /l/ as word-initial. However, these cues disagreed in the case of word-final /l/, since glottalization was always absent in our stimuli materials. While velarization would indicate a word-final /l/, the absence of glottalization in the first vowel of the target word would be a cue for word-initial /l/. Crucially, participants did not always segment the word pairs as /l/-initial in this discrepant case, indicating some amount of sensitivity to the extent of velarization.

Another possibility is that participants indeed only used a lesser degree of velarization as a cue but not a stronger extent of velarization. Participants may have processed velarized /l/ as potentially ambisyllabic, e.g., “kneel leg”, whereas they processed unvelarized /l/ as more likely to be word-initial. Similarly, Nakatani and Dukes (1977) found that their participants sometimes perceived stimuli with spliced in /l/ from the other position, such as “we’ll own”, as having an /l/ both word-finally and word-initially, such as “we’ll loan”. Alternately, it is possible that there is more speaker variability, both within and between, in the extent of /l/ velarization. If this is the case, the

velarization of /l/ would be an unreliable cue, and therefore not particularly useful to the language system in ruling out lexical competitors. One other potential explanation is based on how we determined the point of disambiguation for each condition. Our analyses relied on averaged points of disambiguation for each of the word-final /l/ and word-initial /l/ conditions, and this may have obscured some time-course related effects. Additionally, previous work has suggested that velarization is not a binary feature, but rather a gradient one (Sproat & Fujimara, 1993). Future analyses will aim to account for individual points of disambiguation for each critical item set, and to examine the extent to which velarization predicts segmentation in online and offline tasks.

As discussed in the introduction, we propose two further experiments using German-accented English stimuli instead of native English. Both a categorization and eye-tracking task will be conducted, using the same stimuli as the present experiments but spoken in German-accented English, to examine differences in cue use for word segmentation. Given that most native German speakers have no velarized lateral approximant in their native phoneme inventory, we expect the selected German speaker to produce light /l/ in both coda and onset position, or at least less velarization than for native English. Using acoustic analyses, we will demonstrate that native German speakers do not fully apply this position-sensitive English phonological rule in German-accented English speech. Ideally, these German-accented English stimuli will be matched in duration with the native English stimuli and will also avoid glottalization of vowel-initial targets, so that we may directly compare results using the two speakers. As participants were sensitive to the absence of velarization (and glottalization) as a cue for word-initial /l/, and glottalization did not completely outweigh velarization as a cue for

word-final /l/, our next aim is to examine whether native English listeners missegment word sequences /l/ in coda position of the first word in the sequence when the /l/ is not velarized in German-accented speech. If so, they would show a bias toward interpreting word-final, non-velarized /l/ as word-initial to an even greater extent than was seen in the present experiment, resulting in missegmentation rather than just confusion. Using a blocked design with counterbalanced block order, we will additionally investigate how listeners learn about this accent feature over time and repeated exposure. Recall that the competitor onscreen is not the segmentation alternate itself, but a phonological competitor of that alternate option. Thus, after the point of disambiguation, it becomes apparent what the intended item is, effectively giving feedback to participants about the intended item. Perhaps more importantly, we examine the immediate effects on competition due to this accent feature and longer-term effects of repeated exposure to this accent feature. If native English speakers can learn about the accent feature and override the native tendency to treat [l] as word-initial, they should demonstrate an improvement in distinguishing targets from lexical competitors, and hence in their online segmentation ability.

This study suggests that native English listeners can use velarization of /l/ early in lexical access as a word segmentation cue. Future experiments will examine the effect of foreign-accented speech in which the phonological rule underlying velarization of word-final /l/ is not applied on the segmentation of the speech stream.

CHAPTER IV

GENERAL DISCUSSION

In the experiments conducted for this thesis, we examined processing and learning about phonological rule misapplications in foreign-accented speech and also established a basis for future experiments in this area. In Experiment 1, we examined whether training on phoneme-to-grapheme correspondence in a novel foreign language, which imparts implicit knowledge about a phonological rule, could assist listeners in understanding foreign-accented speech to a language in which that rule is not usually applied. Results showed that this information was not used cross-linguistically. In Experiment 2, we investigated the usage of a word segmentation cue and found that listeners can use the velarization of /l/ in online word segmentation. Proposed follow-up experiments will examine how listeners interpret and learn about foreign-accented speech in which this cue is missing due to its absence in the native language of the speaker. In combination, these experiments will provide insight into the flexibility of phonological knowledge pertaining to foreign-accented speech: whether phonological knowledge about an accent feature can be transferred cross-linguistically, and how phonological knowledge about an accent feature affects the time-course of word segmentation and lexical competition with repeated exposure. A foreign accent may at first be a hard thing to hear, but as listeners you'll learn—but perhaps not from everything.

APPENDIX

LANGUAGE QUESTIONNAIRE

Language Questionnaire

Subject Code: _____

Date: _____

Below are some questions about your language background. Please circle yes or no where applicable.

1. Have you ever learned a foreign language? **Yes** **No**

If yes, please fill in the following sections for each language:

a. Language: _____

For how long did you learn it? _____

When did you start?

How did you learn it?

In school At home Living in another country

Other – please specify: _____

Rate how well you know the language (*circle one number*):

Not Good— 1 2 3 4 5 —Very Good

[Provided space to list 2 more languages.]

If you need more space for this section, please let the experimenter know.

2. Have you ever traveled to any of the following countries? **Yes** **No**

Germany, Austria, Switzerland, Liechtenstein, Belgium, Luxembourg

(Please circle yes if you have traveled to one or more of these countries)

If yes, please fill out the following sections for each country in the list that you have traveled to:

a. Country: _____

How many times did you visit that country?

For about how much time in total did you visit that country?

[Provided space to list 2 more countries.]

If you need more space for this section, please let the experimenter know.

- | | | |
|---------------------------------------------------------------------------------------------|------------|-----------|
| 3. Do you speak any German? | Yes | No |
| 4. Have you taken any German in school? | Yes | No |
| 5. Are you interested in German culture or language? | Yes | No |
| 6. If you have never visited Germany, do you ever want to? | Yes | No |
| 7. Do you know anyone (in your academic, work, social life, etc.) who speaks German at all? | Yes | No |

If yes:

- | | | |
|----------------------------------------------------------------------------------------------------|------------|-----------|
| a. Do any of these people speak German as a first language? | Yes | No |
| b. When they speak in English, do you think that they sound like a fluent, native English speaker? | Yes | No |
| c. To the best of your knowledge, have you ever heard anyone speak in a German accent in person? | Yes | No |

- | | | | | | | | |
|------------------------------------------------------------------------------------------|--------------|---------------|------------------|--------------|---|---|-------|
| 8. How often do you watch subtitled German-language TV or movies?
(Please circle one) | Never | Rarely | Sometimes | Often | | | |
| 9. How often do you listen to German-language music?
(Please circle one) | Never | Rarely | Sometimes | Often | | | |
| 10. In general, how well do you <i>like</i> to learn new languages? (Circle one number) | Dislike— | 1 | 2 | 3 | 4 | 5 | —Like |
| 11. In general, how <i>easy</i> do you find learning new languages? (Circle one number) | Difficult— | 1 | 2 | 3 | 4 | 5 | —Easy |

REFERENCES

- Altenberg, E. P., & Vago, R. M. (1983). Theoretical implications of an error analysis of second language phonology production. *Language Learning*, 33(4), 427-447.
- Brainard, D. H. (1997) The Psychophysics Toolbox. *Spatial Vision*, 10, 433-436.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. Available from <http://CRAN.R-project.org/package=lme4>
- Barr, D. J. (2008). Analyzing “visual world” eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457–474. doi:10.1016/j.jml.2007.09.002
- Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science*, 14(6), 592-597.
- Boula de Mareüil, P., & Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent. *Phonetica*, 63(4), 247–67. doi:10.1159/000097308
- Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707-729.
- Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, 116, 3647.
- Davies, Mark. (2008) *The Corpus of Contemporary American English: 450 million words, 1990-present*. Available online at <http://corpus.byu.edu/coca/>.
- Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, 24(4), 423-444.
- Eisner, F., Melinger, A., & Weber, A. (2013). Constraints on the Transfer of Perceptual Learning in Accented Speech. *Frontiers in Psychology*, 4, 1–9. doi:10.3389/fpsyg.2013.00148

- Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, *15*, 67–83.
- Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5 Pt 1), 3125–34.
- Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech*, *24*(2), 125–146. doi:10.1177/002383098102400202
- Flege, J. E., Schirru, C., & MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*(4), 467-491.
- Giles, S. B., & Moll, K. L. (1975). Cinefluorographic study of selected allophones of English /l/. *Phonetica*, *31*(3-4), 206–27.
- Hallett, P. E. (1986). Eye movements. In K. Boff, L. Kaufman & J. Thomas (Eds.), *Handbook of perception and human performance*. New York: Wiley.
- Hanulíková, A., & Weber, A. (2012). Sink positive: Linguistic experience with th substitutions influences nonnative word recognition. *Attention, Perception, & Psychophysics*, *74*(3), 613-629.
- Howell, P., & Dworzynski, K. (2001). Strength of German accent under altered auditory feedback. *Perception & Psychophysics*, *63*(3), 501–13.
- Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, *19*(4), 332-338.
- Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, *13*(2), 262-268.
- Lane, H. (1963). Foreign accent and speech distortion. *The Journal of the Acoustical Society of America*, *35*(4), 451. doi:10.1121/1.1918501
- Lehiste, I. (1964). Acoustical characteristics of selected English consonants. *Indiana university research center in anthropology, folklore, and linguistics*, *34*, 10–50.

- Levi, S. V., Winters, S. J., & Pisoni, D. B. (2011). Effects of cross-language voice training on speech perception: whose familiar voices are more intelligible? *The Journal of the Acoustical Society of America*, *130*(6), 4053–62. doi:10.1121/1.3651816
- Lewis, M. P., Simons, G.F., & Fennig, C.D. (Eds.). (2013). *Ethnologue: Languages of the World*, (17th ed.). Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Marslen-Wilson, W.D. (1990). Activation, competition, and frequency in lexical access. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. Cambridge, MA: MIT Press.
- McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*, *60*(5), 661–71. doi:10.1080/17470210601183890
- Munro, M. (1995). Nonsegmental factors in foreign accent. *Studies in Second Language Acquisition*, *17*, 17-34. doi:10.1017/S0272263100013735.
- Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*(3), 289-306.
- Munro, M. J., Flege, J. E., & MacKay, I. R. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics*, *17*, 313-334.
- Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *The Journal of the Acoustical Society of America*, *62*(3), 714–719.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. doi:10.1016/S0010-0285(03)00006-9
- Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *The journal of the Acoustical Society of America*, *54*(5), 1235-1247.
- Oxley, J., Roussel, N., & Buckingham, H. (2007). Contextual variability in American English dark-l. *Clinical Linguistics & Phonetics*, *21*(7), 523–42. doi:10.1080/02699200701356485

- Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29(2), 191-215.
- Port, R., & Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics*, 17(4), 257-282.
- R Core Team (2014). R: A language and environment for statistical computing. [computer software] R Foundation for Statistical Computing, Vienna, Austria. Available from <http://www.R-project.org/>
- Recasens, D., Fontdevila, J., & Dolors Pallarès, M. (1995). Velarization degree and coarticulatory resistance for /l/ in Catalan and German. *Journal of Phonetics*, 23(1-2), 37-52.
- Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: lexical stress drives eye movements immediately. *Quarterly Journal of Experimental Psychology*, 63(4), 772-83. doi:10.1080/17470210903104412
- Reinisch, E., Jesse, A., & McQueen, J. M. (2011). Speaking rate from proximal and distal contexts is used during word segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), 978-96.
- Reinisch, E., & Weber, A. (2012). Adapting to suprasegmental lexical stress errors in foreign-accented speech. *The Journal of the Acoustical Society of America*, 132(2), 1165-76. doi:10.1121/1.4730884
- Rogers, C. L., Dalby, J., & Nishi, K. (2004). Effects of noise and proficiency on intelligibility of Chinese-accented English. *Language and Speech*, 47(2), 139-154. doi:10.1177/00238309040470020201
- Sidasas, S. K., Alexander, J. E., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, 125, 3306.
- Simon, E. (2010). Phonological transfer of voicing and devoicing rules: Evidence from L1 Dutch and L2 English conversational speech. *Language Sciences*, 32, 63-86.

- Sjerps, M. J., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 36(1), 195–211. doi:10.1037/a0016803
- Smith, B. L., Hayes-Harb, R., Bruss, M., & Harker, A. (2009). Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics*, 37(3), 257-275.
- Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, 21, 291–311.
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18(6), 645-659.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632-1634.
- Wester, F., Gilbers, D., & Lowie, W. (2007). Substitution of dental fricatives in English by Dutch L2 speakers. *Language Sciences*, 29(2), 477-491.
- Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics*, 15(4), 399-420.