Word Learning Under Accent Variability in Monolinguals and Bilinguals

2010-2011 Honors Project

Kristin Muench
Language Acquisition and Sound Recognition Lab
Department of Cognitive Science, University of California - San Diego
June 2011

Keywords: perceptual learning, speech perception, artificial lexicon, word learning, eye tracking, accent, phonemic flexibility

In accent perception, individuals frequently and easily learn to map words with dissimilar acoustic characteristics onto the same word concept while simultaneously maintaining rigid word boundaries between similar-sounding utterances. Despite these learned perceptual flexibilities, individuals maintain stable linguistic precepts of an unfamiliar accent that can be transferred onto novel incidents of that accent. No existing cognitive mechanism adequately explains this human faculty for understanding and encoding unfamiliar accented language. This experiment provides evidence that unfamiliar accented language encoding may be influenced by an English-Spanish bilingual or an English monolingual language background. Furthermore, the results indicate that monolinguals and bilinguals may learn and encode novel nouns differently when the individual is presented with two accented inputs. Gaining insights into human variability in the linguistic encoding process may facilitate communication and benefit verbal computer interfaces seeking higher-quality user interaction.

1. Introduction

It is unclear how humans understand language within the context of different accents. Individuals routinely map spoken sounds with richly different acoustic information onto the same phoneme, and interpret acoustically similar speech as different phonemes depending on the perceived accent context.

Listeners are sensitive to systematic linguistic variation (Evans and Iverson, 2004); however, systematic variation is abundant in real-world accent learning. Phonemes pronounced in different accents have dramatically different phonological qualities (Hansen, 1995). Different speakers may project dissimilar absolute formant frequencies or formant ranges for the same vowel (Rosen, 1999; Broadbent, 1957). Despite these and other variables that could affect the intelligibility of utterances, humans are able to maintain stable linguistic precepts within a language (Schmid and Yeni-Komshian, 1999).

Certain studies suggest that listeners are excellent at adapting to novel real-world and experimentally created accents. American English speakers could learn to understand spoken Spanish-accented English despite its regular phoneme alterations, even generalizing that accent to novel words and sentences, after only brief exposure to six familiar and unfamiliar talkers (Sidaras, 2009; Clarke and Garrett, 2004). Listeners were also able to adapt to accents that were created for experimental purposes by a front-vowel replacement scheme (Maye, 2008). A mechanism may be at work that applies not just to learned accents but to patterned speech and, possibly, acoustical phenomena in general.

Bilingual speakers may have an advantage over monolingual speakers in perceiving novel accents or in switching accent perception. An individual with a diverse language background may be more likely to detect subtle pronunciation cues that could help them to differentiate between words and, therefore, facilitate accent switching (Barry, 1974). Bilingual speakers do not appear to place vowel category boundaries differently depending on the language context in which it is said (Bürki-Cohen and Grosjean, 1989); however, the language-mixed vocabulary of polyglot aphasics suggests that bilingual learners mix the phonetic systems of their two languages, which might additionally extend to mixing phonemes (Grosjean, 1985). This is further supported by the notion that second language learners may use their first phonemic system when learning a second phonemic system (Pajak and Levy, 2011).

It is nevertheless unclear how bilingualism might affect an individual's ability to learn novel words in a multi-accent context. If bilinguals are privy to a greater variety of acoustic information when confronted with a new accent, then they may take different approach than monolinguals when reconciling these new sounds to words that they already know. Certain types of bilinguals, including English-Spanish bilinguals, may be more flexible when mapping alternate pronunciations of words onto those words due to their proficiency with two sets of vowel spaces with different boundaries (Flege, 1994).

Alternatively, it is possible that bilingualism has no impact on novel word learning in multi-accent contexts, and so the cognitive features that accompany bilingualism do not impact how words are learned and retained in a multi-accent context. I nevertheless expect bilinguals to perform differently than monolinguals in word learning tasks because they seem to possess other traits that would impact the pattern recognition essential to novel language comprehension, such as cognitive control (Bialystok, 2004).

This study observes the performance of English-Spanish bilinguals relative to monolinguals in time to learning and recall of words learned through speech inputs in different fabricated accents with regularly altered vowels. The data gathered through this experiment will provide insights into the cognitive techniques that impact word learning and linguistic pattern recognition in general, as well as how accents are perceived. I expect that all participants will learn the name of the object in two accents and learn to associate the speaker with the accent (Eisner and McQueen, 2005).

Former studies (Bradlowe and Bent, 2008) have primarily investigated adaptation to a novel accent given knowledge of the participant's existing accent. It is worth investigating whether or not accent adaptation can occur in either of two accents that have just been learned simultaneously through a training period. We expect that subjects learning words from multiple accented inputs will take more training blocks to reach a threshold level of accuracy. However, I do not expect words learned under conflicting accented inputs to be recalled with any less accuracy than words learned under a single accent once threshold identification accuracy is reached for both, unless the two conflicting accented inputs contain different words with similar acoustic properties. I would expect, in that case, to see a significant difference in word identification performance between the one accent and two accent conditions.

II. Method

2.1 Participants

Self-reported English monolingual (*n* = 21) and English-Spanish bilingual (*n* = 21) participants between the ages of 18 and 30 were recruited from the Psychology and Cognitive Science departments of the University of California – San Diego. Participants were compensated with class credit for their participation. Monolingual participants used English almost exclusively and fewer than two years of a foreign language. Bilingual participants reported proficiency on both a quantitative and qualitative language background questionnaire (Gollan, in press). The bilinguals primarily described themselves as Mexican Spanish speakers (n= 7), speakers who had learned Spanish in a Southern Californian school (n=4), and Spanish speakers from unspecified cultural backgrounds who regularly interacted with spoken Mexican Spanish (n=9). Almost all bilinguals (n=16) learned Spanish before the age of 3.

2.2 Stimuli

Auditory stimuli were recorded by a male and a female speaker who had lived continuously in southern California since the age of 9. Participant voices were recorded for the auditory stimuli using a soundproof recording room in the Center for Research in Language at UCSD.

The task required participants to learn monosyllabic novel noun names following the consonant-vowel-consonant structure for sixteen unfamiliar object images. Multiple conditions existed in which the nouns associated with stimuli object were switch to correct for selection biases. Four sets of four objects shared an initial consonant. The object associated with the noun was always discernible from the initial consonant and the final consonant.

For each experimental condition, either one, both, or neither speaker referred to the novel noun in an accent that varied only by lowering the front vowel (Fig. 1). Maye, Aslin, and Tanenhaus (2008) previously used this artificial accent paradigm with real English words. The words were designed such that certain noun pronunciations could be confused with each other across accents if no final consonant or talker-specific accent information were present. In one-accent condition, the participant would be trained and tested only on words from one accent (A or B); in the two-accent condition, the participant would be trained on tested on words from both lists.

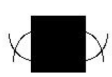
Object (V-initial) Family	Noun (Accent A Pronunciation)	Noun (Accent B Pronunciation)
	Veeg	Vihg
	Vihf	Vehf
	Vehd	Vad
	Vab	Vahb

Figure 1: *Four novel noun stimuli and associated object targets.*

2.3 Procedure and Equipment

The experiment was administered in a low-noise room containing only the testing equipment. Participants were seated in a comfortable chair approximately two feet from the computer screen on which visual stimuli appeared. Audio stimuli were administered at a standard 70 dB amplitude through high quality Sennheiser HD 280 Pro headphones. Participants were allowed adjust the volume of the stimuli to a comfortable level. Participants interacted with the screen both with a standard keyboard and a mouse.

Experimental stimuli were presented using MATLAB and the PsychToolBox 3 (Brainard & Pelli, 1997) and the Eyelink Toolbox (Cornelissen, Peters, & Palmer, 2002). Participants' eye movements were recorded during trials using the remote mode of an Eyelink 1000 at 500 Hz. The experimental presentation computer sent time-stamped messages to the eye tracking computer for later data analysis.

During the experiment itself, participants first completed a questionnaire about their linguistic background immediately prior to participating in the experiment. Previously published experiments also used the same questionnaire to gather data on participant language background (Creel, 2009; Dunn, 2009). Participants who described proficiencies in languages other than Spanish or English in this section were excluded from the analyzed data set.

Participants then proceeded to the main experiment. This experiment presented novel vocabulary items, in both a *training phase* and a *testing phase*. During training, participants were asked to learn the names of 16 different objects through trial and error by clicking on the corresponding noun image in a two-alternative forced choice task. Training phase targets were always paired with competitors that shared neither an initial consonant nor a front vowel in either accented pronunciation. After responding to the auditory stimuli, participants were shown the correct answer, regardless of whether or not they had selected the stimuli correctly. Individual trials within a block were forced-choice selection tasks, in which the participant selected one of two images in response to the stimuli prompt "Choose the…" followed by a novel monosyllabic noun containing a front vowel. Participants only progressed to the test trial once they had reached a 90% correct selection threshold across a training block (134 trials). The training trials were quasi-randomly presented for each repetition of the training block.

In the testing block (192 trials), participants again heard a word on each trial and selected a particular shape. However, no feedback was given for either correct or incorrect answers. During the testing block, participants either encountered same-onset trials (SOTs) or different-onset trials (DOTs). In SOTs, the two pictured objects were labeled by words with the same consonant. For *difficult SOTs,* the competitor object was a different word in the other talker's accent that used the same consonant-vowel (CV) combination. This similarity could be potentially confusing for participants in the two-accent condition. For *easy SOTs*, the competitor object did *not* use the same CV in the other talker's accent. In DOTs, the two words began with different consonants. For *difficult SOTs*, the competitor object was a different word in the other talker's accent that used the same vowel only. *Easy SOT* competitors did not share any initial consonants, final consonants, or front vowels with the targets for those trials (Fig. 2).

Following the experiment, participants performed a short computer task that verified their ability to distinguish discrete phonemic differences in front vowels.

	Onscreen Target		***Onscreen Competitor***	
	Accent A	**Accent B**	**Accent A**	**Accent B**
Easy DOT	Veeg	Vihg	Dehg	Dag
Hard DOT	Veeg	Vihg	Zihb	Zehb
Easy SOT	Veeg	Vihg	Vehd	Vad
Hard SOT	Veeg	Vihg	Vihf	Vehf

Figure 2: *Examples of noun names for the onscreen target and competitor images for all testing trial types. Participants in the one accent condition would only be familiar with words in Accent A or Accent B; participants in the two accent condition would be familiar with words in both Accent A and Accent B.*

III. Results

Training data were analyzed in R for the number of training quarter-blocks required to reach a threshold percentage of correct selections for both monolingual and bilingual participants. Quarter-blocks were analyzed rather than full blocks as a measure that more accurately pinpointed when participants reached a 90% threshold accuracy, which was used as an indicator that participants had learned the object names. A mixed ANOVA indicated that the number of training accents had a significant impact on the number of training blocks (p=0.006), but only monolinguals in the one-accent and two-accent conditions experienced a statistically significant difference according to Welsh's t-test (p=0.017) (Fig. 3).

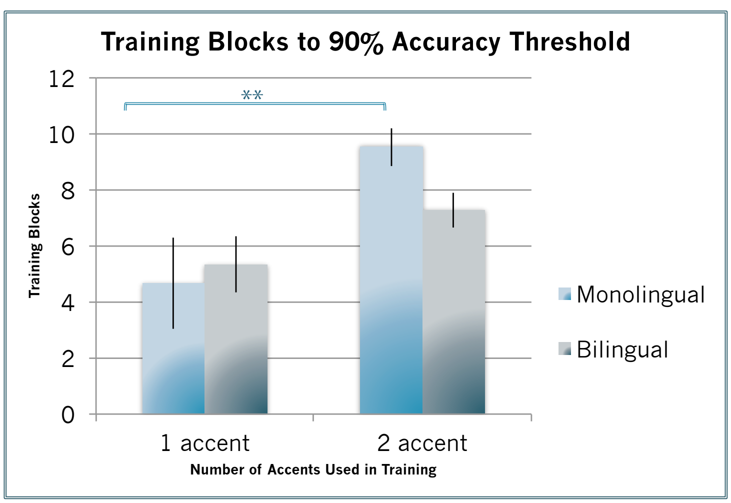


Figure 3: *Number of training blocks needed to reach threshold accuracy for both monolingual and bilingual participants in one- and two accent conditions. The increase in trial blocks between accent conditions was significant for monolinguals (p=0.0172), but not for bilinguals (p=0.1357).*

Testing data were analyzed across all participants groups for the number of correct selections during a testing block. When all types of DOT and SOT trials were considered together within their respective categorizations and the one-accent and two-accent conditions in both monolinguals, a mixed ANOVA indicated for both monolinguals and bilinguals that the type of trial administered affected target identification accuracy. However, it also indicated a significant interaction of accent and type (p=0.001) such that monolinguals in the two-accent condition did markedly worse than would otherwise be expected on SOT trials (Fig. 4). Bilinguals did not experience such an interaction (p=0.746). Statistically significant differences were observed between trial types for monolinguals

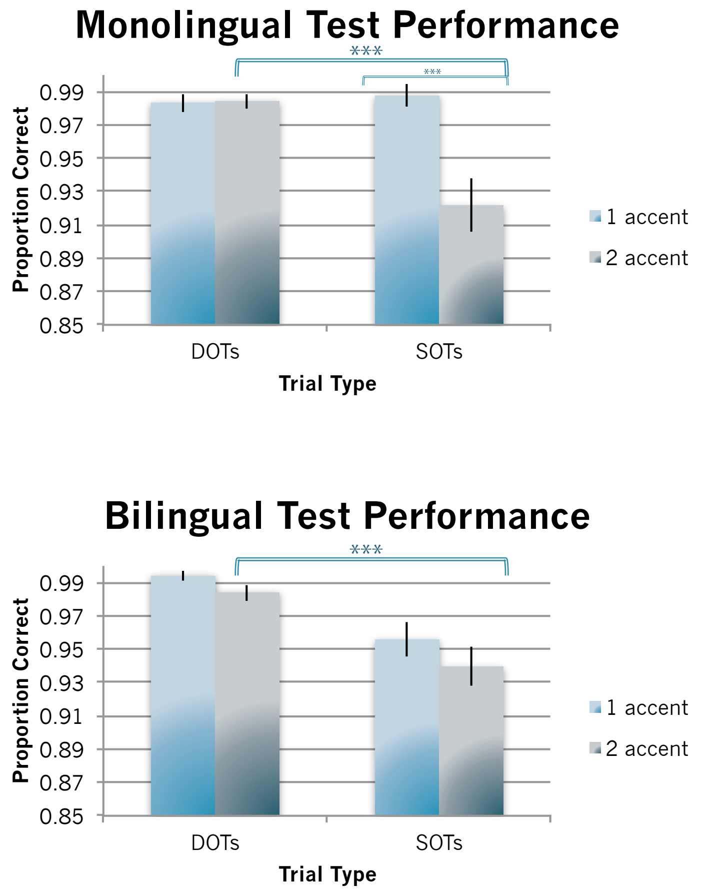
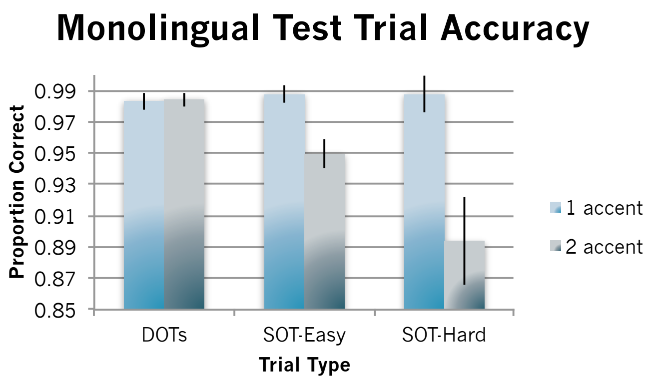


Figure 4: *Proportion of correct selections across testing block for Different-Onset Trials (DOTs) and Same-Onset Trials (SOTs) for both monolinguals and bilinguals in both 1- and 2 accent conditions. Statistical significance was observed through a Welsh's t-test between trial types in 2 accent monolinguals (p=0.00054) and bilinguals (p=0.0006), as well as between accent conditions for monolingual SOTs (p=0.00084).*

A further examination of monolingual accuracy across accent conditions in specific types of SOT trials indicated that both easy and hard SOT trials contributed to the accent condition-trial type interaction effect in monolinguals. The effect seems to be carried to a greater extent by SOT-Hard trials, although there was a statistically significant difference between the one-accent and two-accent conditions for both SOT-Easy (p=0.004) and SOT-Hard(p=0.011) trials. Bilinguals only experienced a statistically significant difference across accent conditions in the SOT-Hard trials, although the baseline for the bilingual participants in the one-accent condition for that trial type was also lower (Figure 5).

III. Discussion



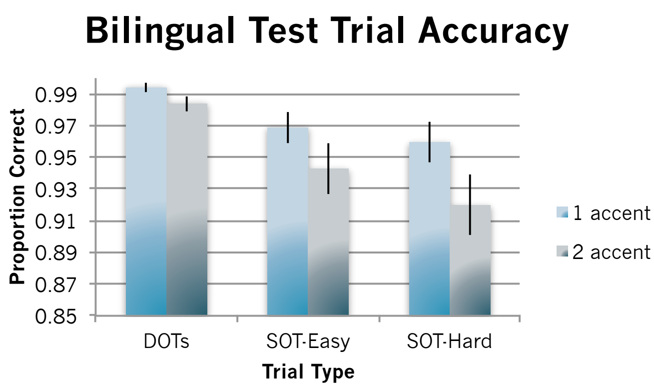


Figure 5: *Proportion of correct selections across the testing block for specific SOT trial types for both monolinguals and bilinguals in both 1- and 2 accent conditions. Monolinguals performed significantly differently across accent conditions on SOT-Easy (p=0.00036) and SOT-Hard (p=0.0114) trials. Bilinguals experiences a significant difference in test trial accuracy in the two accent condition between SOT-Easy and SOT-Hard trials (p= 0.0073).*

The results suggest that English-Spanish bilinguals learn words differently under multi-accented inputs than English monolinguals.

All participants performed exceedingly well when tested for the ability to distinguish between discrete phonemic differences in front vowels. Therefore, all participants seemed capable of perceiving and identifying the differences between the manipulated vowels.

Eyetracking data suggests that all participants used talker-specific vowel information to associate the novel noun with the visual stimulus. The stimuli words were designed such that the target could be identified from the initial consonant and the vowel alone for all but SOT-Hard trials. In that case, the participant would need to rely on knowledge of the talker's accent in order to determine the target without hearing the final consonant. The fixation locations for a participant were binned across 50 millisecond blocks of time within a trial. The probability that a participant would look at the target object versus the competitor across that time chunk (i.e., Target Advantage) was plotted for different trial types. A more positive Target Advantage indicates that a participant was more likely to be gazing at the target at that time point for that trial type, and the Target Advantage was expected to increase throughout the trial as the participant gained more acoustic information about the word. On average, across all test trials, participants should have fixated on the target after hearing the final consonant 600 milliseconds into the word. A Welsh's t-test for each trial type in each group determined that participants already had a positive Target Advantage significantly different than zero at that timepoint for every trial type, including the Cohort-Hard trials (Fig. 7). Therefore, participants were more likely to look at the target before hearing the final consonant, and thus seem to use knowledge of an accent associated with a speaker to select a target after hearing a prompt.

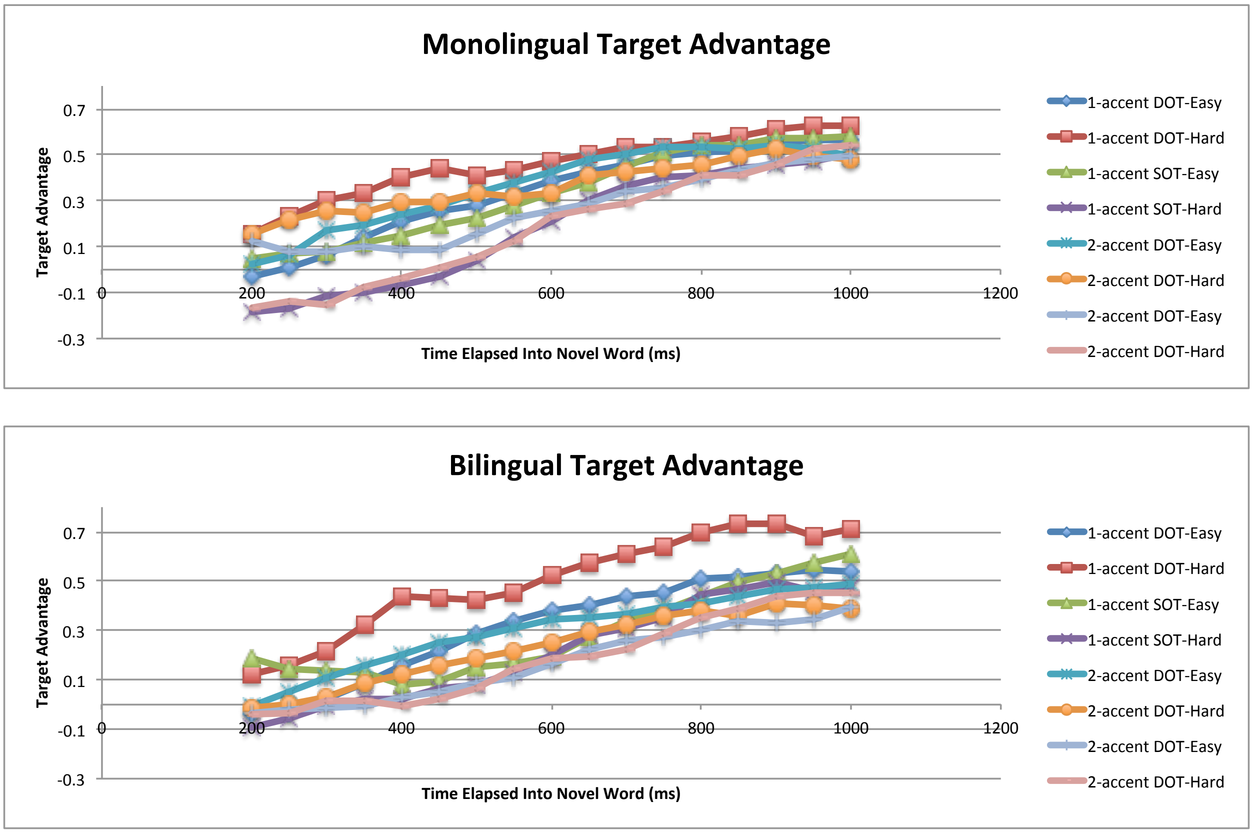


Figure 7: *Target Advantage (i.e. likelihood that a participant is looking towards the target), calculated from monolingual and bilingual eye tracking data. Participants relying on final consonant information would not experience a Target Advantage significantly above 0 until after the 600 millisecond time-point.*

Although the results from the learning period do not prove a bilingual advantage in word learning, they do suggest that monolingual speakers may have a more difficult time learning words given two accented inputs relative to learning words from only one accented input. The threshold of 90% accuracy was established as a point at which participants could be said to have learned the object names, and therefore the number of blocks to 90% accuracy may be interpreted as the number of blocks to learning. Bilinguals in the one-accent and the two-accent conditions, on the other hand, do not experience a dramatic difference in the number of quarter-blocks to learning. This result suggests that there may be some difference in the way bilinguals and monolinguals perceive and process novel words.

In an examination of test trial accuracy, monolinguals were significantly less accurate in cohort trials when they had learned novel words in multiple accents. With the exception of monolingual SOT trials, the number of training accents did not produce a statistically significant difference within any trial type for any group of participants. This contrasts markedly with the cohort trial performance of monolinguals, which was greatly influenced by the number of accents used during the training. Although both monolinguals and bilinguals experienced a significant drop in accuracy on these more difficult trials in the one accent condition, two-accent condition monolinguals performed even worse than would be expected for SOT trials. Bilinguals seemed to be rescued from this effect; they were no less accurate in the two-accent condition than the one-accent condition for cohort trials.

Within two-accent condition monolinguals, participants were the least accurate on SOT-Hard trials, in which the greatest potential for confusion of a word across accents existed. Participant target selection accuracy was only moderately impaired when the competitor shared an initial consonant with the target object. English monolinguals experience a drop in cohort-hard novel word identification trials that is not observed to such an appreciable extent in bilinguals when they must learn the novel words through two conflicting accents. This further underscores the finding that monolinguals and bilinguals process novel accented words differently.

The fact that English monolinguals confuse similar-sounding words despite associating each talker with an accent suggests that they may encode words in each accent in such a way that significant interconfusion can occur between their linguistic sets.

Bilinguals, by seemingly evading this cross-accent confusion, seem to be inclined to keep the representations of the word separate. They may be better at associating the accent with the speaker, and are thus suffer less object-name confusion. It is possible that English-Spanish bilingualism increases the flexibility of vowel category boundaries, thus making it easier to map words with slightly different vowels onto the same word representation. A bilingual may, as a result, have more permissive word boundaries and may have an easier time learning to map a wider range of vowel-shifted words onto the original target.

Further analysis of the rich eye tracking dataset assembled throughout this experiment may elucidate what mechanism motivates this difference between English monolingual and English-Spanish bilingual speakers. Future experiments building on this work may use this paradigm to explore how monolingual and bilingual participants learn words from accented inputs in which one accent's vowels are irregularly shifted relative to the other accent. Such a study might demonstrate the limits of English-Spanish bilingual phonemic permissiveness and provide support for bilingual phoneme flexibility. Armed with an improved understanding of how individuals vary in their extraction and processing patterns from complicated speech environments, individuals may begin to tailor their linguistic interactions to maximize communicative efficacy in an increasingly global society.

1.Alexander, Jessica E D, Nygaard, Lynne C (2009), Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, 125(5): 3306-3316.

2.Barry, W J (1974), Language background and the perception of foreign accent. *Journal of Phonetics.*

3.Bialystok, Ellen (2004) Bilingualism, Aging, and Cognitive Control: Evidence From the Simons Task. *Psychology and Aging*, 19(2):290-303

4.Bradlow, Ann R, and Bent, Tessa (2008), Perceptual adaptation to non-native speech. *Cognition*, 106(2): 707-729.

5.Brainard, D. H. (1997), The Psychophysics Toolbox. *Spatial Vision*, 10:443-446.

6.Broadbent, D E (1957), Information Conveyed by Vowels N recent years a great deal of research. *Applied Psychology*, 29(1): 98-104.

7.Clarke, Constance, and Garrett, Merrill (2004), Rapid adaptation to foreign-accented English. Journal of the Acoustical Society of America, 116(6): 3647-3658

8.Cornelissen, F.W., Peters. E., Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods*, Instruments & Computers, 34, 613-617.

9.Creel, S. C., & Tumlin, M. A. (2009). Talker variability is intrinsic to word representations: Evidence from on-line processing of spoken words. Proceedings of the 31st annual Cognitive Science Conference, Amsterdam, NL.

10.Dunn, Alexandra, and Fox Tree, Jean (2009), A quick, gradient Bilingual Dominance Scale. Language and Cognition, 12(3): 273–28.

11.Evans, Bronwen and Iverson, Paul (2004), Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *The Journal of the Acoustical Society of America*, 115(1): 352–361.

12.Fox, Robert, James Flege, and Murray Munro. "The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis." *Journal of the Acoustical Society of America*. 97.4 (1995)Gollan, Tamar (in press), Assessing Language Dominance: Multilingual Naming Test.

13.Grosjean, François (1985), Polyglot aphasics and language mixing: A comment on Perecman. *Brain and Language*, Vol 26(2): 349-355.

14.Hansen, John. (1995), Foreign accent classification using source generator based prosodic features. *Electrical Engineering*: 836-839.

15.Judith Bürki-Cohen, François Grosjean2 and Joanne L. Miller (1989), Base-Language Effects on Word Identification in Bilingual Speech: Evidence from Categorical Perception Experiments. *Language and Speech* 32(4): 355-371.

16.Maye, J., Aslin, R. N. and Tanenhaus, M. K. (2008), The Weckud Wetch of the Wast: Lexical Adaptation to a Novel Accent. *Cognitive Science*, 32: 543–562.

17.McQueen, James M, Cutler, Anne, Norris, Dennis (2006), Phonological Abstraction in the Mental Lexicon. *Cognitive Science*, 30: 1113-1126.

18. Pajak, Bozena and Levy, Roger (2011), Phonological Generalization from Distributional Evidence. To appear in *Proceedings of the 33rd Annual Meeting of the Cognitive Science Conference*.
19. Rosen, Stuart. (1999), Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. *The Journal of the Acoustical Society of America*, 106 (6): 3629-3636.
20. Schmid, Peggy and Yemi-Komshian, Grace (1999) The Effects of Speaker Accent and Target Predictability on Perception of Mispronunciations. *Journal of Speech, Language, and Hearing Research* Vol.42 56-64
21. Sidaras, Sabrina K, Eisner, Frank, McQueen, James M (2005), The specificity of perceptual learning in speech processing. *Perception & psychophysics*, 67(2): 224-238
22. Sidaras, S.K., Alexander, J.E.D., & Nygaard, L.C. (2009), Perceptual learning and systematic variation in Spanish-accented speech. Journal of the Acoustical Society of America.