Distribution Agreement

On vocal alignment to American- and Spanish-accented talkers

By

Eva M. Lewandowski
Master of Arts
Psychology

____

Lynne Nygaard
Advisor

____

Laura Namy
Committee Member

____

Harold Gouzoules
Committee Member

Accepted:

____

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

____

Date

On vocal alignment to American- and Spanish-accented talkers

By

Eva M. Lewandowski
B.A., Montclair State University, 2012

Advisor: Lynne Nygaard, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Arts
in Psychology
2014

Abstract

On vocal alignment to American- and Spanish-accented talkers

By Eva M. Lewandowski

Vocal alignment is the tendency to change one's speech productions to more closely match those of another individual. Some theoretical accounts hold that high level, social-motivational goals are the underlying cause of this tendency while others propose that low level, perceptual-motor couplings are the cause. The present study sought to disambiguate these potential mechanisms of vocal alignment, exploiting the social and perceptual qualities of Spanish-accented speech. It was predicted that if vocal alignment was primarily social-motivational, vocal alignment would occur less to Spanish-accented speech. The opposite trend would be observed if alignment was perceptual-motor. Study participants took part in a shadowing task paradigm and heard English word-length utterances produced by four model talkers (2 American English, 2 Spanish). Vocal alignment was assessed by both acoustic measures and listener judgments. Results indicated a significant trend towards vocal alignment. When assessed by acoustic measures, participants seemed to align more to native English speakers, consistent with a social-motivational approach. When assessed by listener judgment, participants aligned more to Spanish-accented speakers. These results suggest that an interaction between low level and high level mechanisms leads to patterns of vocal alignment.

On vocal alignment to American- and Spanish-accented talkers

By

Eva M. Lewandowski
B.A., Montclair State University, 2012

Advisor: Lynne Nygaard, Ph.D.

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Arts
in Psychology
2014

Acknowledgments

I wish to acknowledge SooMhin Park, Sophie Cohn, Sana Bhojwani, and Joy Lee for their assistance with the time-consuming tasks of collecting and coding data, as well as Valentin Lazar for his expertise with the experiment software. I also wish to thank Harold Gouzoules and Laura Namy, members of my faculty advisory committee for their insights and suggestions for how to approach my topic of study. I thank Lynne Nygaard for her supportive and critical feedback at every step of the research process, and for feedback on early versions of this manuscript.

Table of Contents

List of Tables, Figures, and Appendices

On Vocal alignment to American- and Spanish-accented Talkers

Speech is a highly mutable signal that varies due to item, listener, talker, and situational factors (Bradlow & Pisoni, 1999; Goldinger, Pisoni, & Logan, 1991; Martin, Mullennix, Pisoni, & Summers, 1989; Nygaard & Pisoni, 1998; Peterson & Barney, 1952; Roodenrys et al., 2002). Individuals are not only perceptually sensitive to these variations, but often alter their own speech productions to more closely match these variations in the acoustic speech signal (e.g., Bilious & Krauss, 1988; Pardo, 2006). This tendency to match vocal characteristics to another speaker, or *vocal alignment*, can occur to the structure of speech along any linguistic level, from the fine-grained aspects of phonetic pronunciation (Babel & Bulatov, 2011; Babel, 2010; Sancier & Fowler, 1997; Shockley, Sabadini, & Fowler, 2004) to higher level linguistic components such as lexical choice (Branigan et al., 2011) and syntactic structure (Branigan, Pickering, & Cleland, 2000). Although many studies have examined the contexts in which vocal alignment occurs and the factors that alter the degree of vocal alignment (e.g., Pardo, Jay, & Krauss, 2010), the mechanisms underlying vocal alignment remain poorly understood.

Two general categories of explanation for vocal alignment have emerged in the literature. One line of research points to social modulation of vocal alignment. It posits that social-motivational factors, such as the desire to be liked, to establish common identity, or to balance social power are primarily responsible for vocal alignment (Bourhis & Giles, 1977; Giles, Coupland, & Coupland, 1991). Studies grounded in a social account of speech alignment do reveal that social status is a relatively strong predictor of alignment. For example, Giles (1973) studied the influence of perceived social power on accent accommodation. In this study, Bristol-accented participants were interviewed by either a speaker of Received Pronunciation (RP), a high status dialect of British English, or were interviewed by a speaker of Bristol English, a lower status dialect of British English. Participants were more likely to align to the high status dialect than to the lower status dialect. This “accent mobility” illustrates that the tendency to alter the characteristics of one’s dialect based on of an interlocutor’s speech patterns may be influenced by social status. This pattern of status-oriented vocal alignment can be found in multiple interview settings, including television talk shows (Gregory & Webster,1996). Relative social standing has been found to interact with attitudes to determine patterns of alignment. In cases where negative attitudes are present, vocal divergence can occur.

Bourhis and Giles (1977) explored this effect by having an RP speaker interview Welsh-English bilinguals. Subjects tended to converge toward the RP interviewer, until they “overheard” a staged event where the interviewer insulted Welsh language and culture. Participants who identified strongly with their Welsh background reacted by diverging from the RP interviewer and adopting a strong Welsh accent, while participants with a weak Welsh identity continued to align with the RP pronunciation. Because the participants rejected RP production norms when they perceived the interviewer as anti-Welsh, Bourhis and Giles concluded that alignment depends on interlocutor attitudes. More recently, Babel (2010) found that a person’s implicit attitudes toward a group of speakers predicted vocal alignment and divergence better than when attitudes were manipulated by a single, external event (i.e. overhearing an cultural insult). Taken together, social attitudes and group identity, which can be long-standing or recently established, will also influence degree of vocal alignment.

Although speakers do generally exhibit a tendency to align (or diverge), implicit socialization of social roles and behaviors, such as gendered behaviors, can exert an influence on alignment beyond social power and attitudes towards a cultural group. One study that sheds light on the gender differences in alignment found that males and females align and diverge on different speech dimensions (Bilous & Krauss, 1988). Men and women aligned on utterance length, pause length, and number of interruptions, but diverged on frequency of laughter. Even when an interactive task is not used, the effects of gender on alignment behavior persist (Namy, Nygaard, & Sauerteig, 2002). Furthermore, gender differences interact with other social-motivational factors. To examine the role of social power and interlocutor gender, Pardo (2006) asked same sex pairs (i.e., male-male and female-female pairs) to complete a map task wherein an instruction giver provided detailed verbal directions for the instruction receiver to draw a specific path on a map. Male instruction givers demonstrated more vocal alignment to instruction receivers, but female instruction receivers aligned more to instruction givers. These studies support a complex interaction among the effects of social motivational factors on vocal alignment.

In order to explain the diverse effects of social-motivational factors on convergence and divergence, Giles and colleagues (Bourhis & Giles, 1977; Giles, 1973; Giles, Coupland, & Coupland, 1991; Giles, & Ogay, 2007) developed the "Communication Accommodation Theory" (CAT) framework. The driving tenet of CAT posits that talkers modulate their speaking styles to create or reduce social distance. Interlocutors converge on each other's speaking styles to foster liking, accomplish a task, or show social affiliation. Interlocutors can also increase social distance or assert an identity distinct from a disfavored affiliation or set of opinions through vocal divergence. While Giles does not make clear whether this process is under the talker's volitional control or if it is the result of social conditioning, evidence from vocal divergence suggests a volitional control must be involved.

If the assumption is that all vocal alignment is under volitional control because vocal divergence must occur with some conscious involvement, some findings on vocal alignment are difficult to explain. For example, Evans and Iverson (2007) studied dialect change of Northern British-accented adults after a year of attending university in a multi-dialectical community. Despite the finding that the participants did change their productions of regionally distinct vowels (as assessed by *acoustic* measures of vowel formants and duration), the differences were not perceptible to community members. Even though alignment had occurred to the dialect norms of the new community, existing community members did not detect the changes. Similarly, Sancier and Fowler (1997) studied the speech productions of a second language learner (L1=Brazilian Portuguese, L2=English). When this talker had extended stays in the US, the duration of voice onset time (VOT) for stop consonants became longer, consistent with English language norms. The change was noticeable to native speakers of Brazilian Portuguese but was imperceptible to native speakers of Standard American English (SAE). In other words, the bilingual talker made an adjustment toward the SAE norms, but it went unnoticed by SAE listeners.

A CAT approach posits that the primary goal of vocal alignment is to modulate social distance. Talkers in both Evans and Iverson (2007) and Sancier and Fowler (1997) did align to their community's norms, suggesting closeness to the community. If the dialect alignment had been driven by explicit social goals, then the talkers should have modulated their speech to be noticeably more similar to their community members. Since participants seemed to align to the ambient language norms enough as detected by Portuguese listeners or acoustic measures, but not enough for English or multi-dialect community listeners to detect, this accommodation likely occurred without a volitional control.

The social-motivational work is also challenged by a second line of research that focuses on experimental paradigms with minimal social constraints. For example, Goldinger (1998) presented prerecorded, isolated words to participants in a shadowing task and listeners were asked to simply repeat each word aloud. Even in this minimally social environment, shadowers aligned the properties of their speech to the model talker. The presence of vocal alignment in the absence of a conversational or interview setting suggests that social factors alone may not be enough to explain the behavior. Shadowing paradigms have been used to examine alignment to fine-grained acoustic and phonetic characteristics of speech and have typically demonstrated alignment despite the lack of direct interlocutor contact (Brouwer, Mitterer, & Huettig, 2010; Kappes, Baumgärtner, Peschke, & Ziegler, 2009; Mitterer & Ernestus, 2008; Pardo et al., 2013; Sancier & Fowler, 1997; Shockley, Sabadini, & Fowler, 2004). Although social factors seem to modulate vocal alignment, this line of work has emphasized the automaticity of vocal alignment, and holds that implicit mechanisms, such as priming or perceptual-motor couplings, primarily drive alignment (Fowler & Galantucci, 2005; Pickering & Garrod, 2004).

Because vocal alignment can occur without the social dynamics present in conversational settings, a number of alternative accounts have been proposed to explain alignment behavior. As one example, Fowler & Galantucci (2005) propose that vocal alignment is due to the tight coupling between speech perception and speech production. According to this view, the goal of speech perception is to recover a talker's intended *articulatory gestures* (intended speech sounds & vocal tract configurations) from a speech signal (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Because the same gestures are necessary for speech perception and for speech production, the two are seen as specially linked (Liberman & Mattingly, 1985). A listener perceives speech by mapping the speaker's articulatory gesture onto his or her own articulatory system. Simulation of the articulatory form enables a listener to better comprehend speech. Vocal alignment occurs because representing another individual's articulatory gestures alters the listener's subsequent production.

In support of the low-level perception-production link, Adank, Hagoort, & Bekkering (2010) found that imitating a novel Dutch accent led to significant improvement in participants' ability to comprehend words spoken in the novel accent. The improvement in intelligibility from imitation was greater than the gain from other types of training, suggesting that alignment is useful for understanding non-standard speech. However, Adank and colleagues (2010) did not measure spontaneous alignment directly, but rather, encouraged explicit imitation. If vocal alignment were truly the result of speech perception-production links, it would be useful to know if alignment occurred spontaneously and if spontaneous alignment improved intelligibility.

Taken together, the research on vocal alignment supports neither a purely perceptual-motor explanation nor a purely social-motivational explanation of vocal alignment. Studies demonstrate a role for both social and automatic factors in vocal alignment. The question, then, is not which one, but rather to what extent and under which circumstances do low-level (e.g., perceptual-motor) and higher-level (e.g. social-motivational) factors work to determine the degree of vocal alignment. One way to merge these parallel accounts of vocal alignment is to consider speech alignment in the context of other behaviors that can become synchronized during the course of an interaction. Situated in such a way, vocal behavior can be conceptualized as part of a suite of other interactive alignments. Evidence that vocal and non-verbal alignment may be linked comes from studies demonstrating that alignment along some verbal dimensions, such as lexical choice, and alignment to non-verbal behaviors, such as manual gesture and postural sway, occurs simultaneously (Louwerse, Dale, Bard, & Jeuniaux, 2012; Shockley, Baker, Richardson, & Fowler, 2007). Although speech and behavioral synchrony are infrequently compared in a single study, if similar factors cause similar patterns of alignment for speech and non-speech behaviors, this may be taken as evidence that they are part of the same imitative system. Assimilating social-motivational and perceptual-motor factors associated with vocal alignment can be achieved with a general synchrony approach.

While the proposed link between speech perception and speech production (e.g. Fowler & Galantucci, 2005) suggests that vocal alignment could occur because of the specific relationship between perception and production of speech, the findings of

Louwerse and others suggest that speech may not be special. Action understanding of any kind can lead to synchrony of behavior (Chartrand & Bargh, 1999; Louwerse et al., 2012). Moreover, behavioral synchrony may be an implicit system default caused by automatic social priming, thus providing a possible integration for social-motivational and automatic factors in vocal alignment (Van Baaren, Janssen, Chartrand, & Dijksterhuis, 2009). This offers an integration of both the social-motivational and perceptual-motor accounts of vocal alignment. As proposed by the perceptual-motor account, the baseline tendency is toward alignment rather than divergence. This is consistent with some existing theories that propose automatic priming as leading to vocal alignment (Pickering & Garrod, 2004). The mechanism is not a module of speech, but rather one that is susceptible to influences of social-motivational information via priming. By including a role for cognitive supervision, vocal divergence can also be explained.

Some evidence that social information acts implicitly on vocal alignment behavior already exists, which mirrors findings of implicit effects of social information on non-verbal alignment. This evidence comes in the form of a shadowing study on vocal alignment (Namy et al., 2002). The key finding from this experiment was a general trend towards alignment. Importantly, gender had an effect on alignment behavior such that women were significantly more likely to align to male rather than to female speakers. The authors suggest this pattern was observed because of differences in gender socialization, which results in an increased sensitivity to indexical cues in women (Namy et al., 2002; Nygaard & Queen, 2000). This is an example of social information implicitly entering into a shadowing task. If the socialization of gender can penetrate a minimally social context, then perhaps vocal alignment operates in a similar way to types of non-verbal alignment. Moreover, if the effects of gender can operate in a minimally social task, then factors typically considered high-level, such as attitudes towards a speaker group, can be observed as well.

Other types of variations in speaking style that influence attitudes toward a speaker group also have implications for vocal alignment. Babel (2010) found that New Zealanders with negative attitudes toward Australians showed less vocal alignment to a speaker of Australian English than New Zealanders with positive attitudes toward Australians. Babel (2012) used a shadowing paradigm where participants were either shown a "picture" of the talker they were about to hear, or given no visual information. Participants who found the talker attractive tended to have greater rates of vocal alignment. These types of findings suggest that attitudes about other speakers may have a profound influence the extent to which speakers vocally align to an interlocutor's utterance. These effects may be even stronger for more noticeable differences in speaking styles, such as a foreign accent.

Alignment to non-native speech

One way to evaluate how much involvement social-motivational and automatic factors have on the vocal alignment process is by investigating alignment to foreign-accented speech. As an acoustic speech signal, foreign-accented speech is both more variable and often less intelligible than native-accented speech (Bradlow & Bent, 2008; Bradlow & Pisoni, 1999; Lane, 1963; Munro, & Derwing, 1995; Sidaras, Alexander, & Nygaard, 2009; Sumner, 2011). These factors present challenges to listeners' ability to comprehend foreign-accented speech, which opens two possible outcomes under a perceptual-motor account of vocal alignment. First, that vocal alignment would only occur between partners of similar language backgrounds because moving from native to non-native pronunciations is simply too far outside the speaker's native linguistic system and representational structure (Kim, Horton, & Bradlow, 2011). Second, because native and non-native speakers are more different than native and native speakers, more alignment would occur because there was a greater starting disparity between speech realizations. In a study of mixed-language dyads, Kim et al. (2011) explore this possibility with a dyadic task. Their results supported, on average, the former of the two potential outcomes. However, the pattern of alignment and divergence across dyads was inconsistent, even for talkers sharing native language backgrounds (Kim et al., 2011; Van Engen et al., 2010). The authors acknowledged the inconsistent patterns of alignment could be due to native speaker's attitudes towards the speakers with accents, which were not assessed. Some evidence also suggests that native English speakers will align to non-native models in a shadowing task paradigm. If task demands can change the observed patterns of alignment (e.g., explicit imitation versus dyadic task), simple perception-production links are not enough to explain vocal alignment behavior.

As a social index, non-native accents convey a host of information to a native listener, including what the talker's native language is, and social factors that accompany language use (Giles & Ogay, 2007; Gluszek, & Dovidio, 2010). For example, Spanish-accented English tends to be judged with negative stereotypes (McKirnan & Hamayan, 1984; Ryan, 1983; Ryan, Carranza, & Moffie, 1977). If vocal alignment is automatic and functions to facilitate language comprehension the way that an action understanding account would predict, then there should be *more* vocal alignment to Spanish-accented speech than to native-accented speech. If vocal alignment is socially motivated and serves to modulate social distance the way a social motivational account would predict, then there should be *less* vocal alignment to Spanish-accented speech than to native-accented speech.

To date, the studies on alignment toward accented speech are mixed, and none use Spanish-accented English. For example, Kim and colleagues (2011) found that participants in a dyadic task were more likely to vocally align if they shared a language background (native Korean or native English) than partners of different dialect or language backgrounds. More recent work, however, has found a trend for native English speakers to accommodate toward Korean-accented speech in a task similar to a standard shadowing paradigm (Kim, 2011). These mixed findings are likely due to the combination of factors, including the overall intelligibility associated with particular non-native accents and relative familiarity with foreign-accented speech. The current experiment examines Spanish-accented speech in order to investigate the relative contribution of social versus perceptual-motor influence on vocal alignment. Foreign-accented speech in general and Spanish-accented speech in particular has two properties that are useful for examining current accounts of vocal alignment: accented speech is less intelligible than native speech, and accented speech carries negative social attributions.

Logistical challenges would make it difficult to manipulate talker and accent familiarity as part of a conversational task. However, it is possible to increase accent familiarity by coupling a perceptual learning paradigm (Bradlow & Bent, 2008; Bradlow & Pisoni, 1999; Sidaras, Alexander, & Nygaard, 2009) with a shadowing task (Goldinger, 1998). Goldinger's (1998) task design included perceptual training trials for his (native English) shadowers before the shadowing component of the shadowing task. The number of repetitions a shadower would hear of each utterance was manipulated across subjects. Increasing the number of pre-shadowing repetitions generally increased ratings of vocal alignment, but this effect hit ceiling at 4-repetitions. Therefore, in the current experiment, Goldinger's paradigm was modified to include Spanish-accented utterances with a perceptual learning component that served the dual purpose of maximizing the likelihood of vocal alignment and familiarizing individuals with Spanish-accented speech. The shadowing paradigm also allows the social context to be minimized, reducing potential social-motivational effects beyond those evoked by the talker's voices and those brought to by the participant to the testing setting.

Mitterer and Müsseler (2013) argue that a linguistic feature must be made salient in order for it to be the object of alignment. To support their claim, the authors studied the degree of phonetic imitation of regional German dialects. Participants tended to converge toward a non-standard production if variation was present in the stimulus set, but did not change their productions if the stimulus set maintained dialect constancy. For this reason, I included multiple model talkers in the experimental task. Individuals have unique speech characteristics, such as voice, vowel space, and speaking rate, which enables talker identification (Doddington, 1985; Sheffert, Pisoni, Fellowes, & Remez, 2002). Because exposing listeners to multiple talkers invites comparison of the talkers (Bradlow & Bent, 2008; Goldinger, 1998), including multiple talkers in a shadowing task paradigm should allow listeners to hear the most salient voice characteristics of each model talker. Thus, individuals should align to the talkers' speech because idiosyncrasies are made more salient.

Using a shadowing paradigm with Spanish-accented speech stimuli, the current study will explore the mechanisms underlying vocal alignment. Participants heard utterances from multiple model talkers, some of whom were native English speakers and some of whom were native Spanish speakers. To assess alignment, listeners compared the shadowers' baseline, the model's, and the shadower's shadowed utterances (Goldinger, 1998; Namy et al., 2002). While this method is useful for determining global alignment, it does not identify specific dimensions along which alignment has occurred. As such, vocal alignment was also assessed acoustically, through measurements of fundamental frequency (f0) and duration. According to a social-motivational account of vocal alignment, there should be less alignment to Spanish-accented speech than to native-accented speech because of the negative stereotypes often associated with Spanish accents. According to an automatic, action-understanding account, there should be more alignment to Spanish-accented speech than to native-accented speech because foreign accents are more challenging to comprehend.

Methods

Participants

Models. Four talkers were selected from the Speech and Language Perception Lab's speech databases to serve as models for the shadowing task. Two of the selected talkers were male and two were female. One male (EM) and one female (EF) were native speakers of Standard American English. One male (SM) and one female (SF) were native speakers of Mexico City Spanish. Mean age of arrival for the Spanish talkers was 26.42, and mean age at which they began learning English was 16.67. The Spanish models were selected on the basis of their mean accentedness scores, sentence-level intelligibility, and word-level intelligibility (see description below). All model talkers were recruited from the Atlanta area, and reported no hearing and speech disorders.

Shadowers. Thirty-three native English speakers (7 male and 26 female) participated in the shadowing task. All shadowers reported no history of speech or hearing disorders at time of participation. Ten shadowers were excluded on the basis of language background (e.g., reported being a non-native English speaker; 2 shadowers) or due to computer technical error (8 shadowers). The remaining 23 shadowers (6 male and 17 female) were all native English speakers who were unfamiliar with Spanish. Their utterances were used in the subsequent AXB task. Shadowers received partial credit towards an introductory psychology course requirement for their participation.

AXB raters. One hundred and thirty-six native English-speaking raters completed the AXB judgment task. Seven raters' judgments were excluded due to computer technical error. None of the raters had participated in the shadowing task, and raters reported no history of speech or hearing disorders at time of participation. AXB raters received partial credit towards an introductory psychology course requirement for their participation.

Stimuli

For all phases of the experiment, a single word list composed of 72 "hard" CVC (consonant-vowel-consonant) words was used. As defined by Luce and Pisoni (1998), hard words are low frequency and have many high frequency phonological neighbors. Word stimuli for the current study were low frequency ($\bar{x}$=12.22; Kučera and Francis, 1967) with many ($\bar{x}$=282.22) high frequency neighbors. The word list sampled from eleven American English vowel categories, including three diphthongs (/oʊ/, /eɪ/, /aɪ/) and eight monophthongs (/ɒ/, /æ/, /i/, /ɪ/, /ɛ/, /u/, /ʌ/, /ɔ/). Vowel and consonant classes were not represented to an equal extent in the word list, and semantic content was not controlled. The full word list is reported in the Appendix. The four model talkers were recorded reading the list of words using Audacity recording software. Sound files were digitized at 22.050 kHz, amplitude normalized, and segmented into word-length sound files. These utterances were used in the shadowing task and as the comparison stimuli in the AXB judgment task. Shadowers' utterances were recorded and treated identically to model talkers' utterances for use in the AXB task.

Intelligibility and Accentedness Judgments. Spanish model talker selection was based on intelligibility ratings from a transcription-in-noise task of 100 Harvard sentences (IEEE Subcommittee, 1969) and 144 monosyllabic words. The Harvard sentences are monoclausal and contain five key words (e.g., A *gray mare walked before* the *colt*.) The sentences were mixed with white noise at a +10 signal-to-noise ratio for the intelligibility tasks. Separate groups of 10 (native English-speaking) listeners transcribed all sentences and words to determine word-level and sentence-level intelligibility. Baseline intelligibility for the sentences was calculated by assessing the percentage of correct transcriptions of key words. Ten additional English-speaking participants provided an accentedness rating for each Spanish model based on ten sentence-length utterances from each model presented in the clear. Accentedness scores were generated on a 7-point Likert-type scale where 1= "not accented" and 7="very accented." The database of Spanish-accented speech contained accentedness and intelligibility ratings from 12 (6 male and 6 female) Spanish speakers. The male and female with the lowest accentedness score and highest sentence- and word-level intelligibility were selected as the Spanish model talkers for the current study ($SF_{accentedness}$= 3.1, $SF_{sentences}$=89.80, SF_{words}=68.80; $SM_{accentedness}$=2.68, $SM_{sentences}$=90.70, SM_{words}= 60.27).

Procedure

The Emory University Institutional Review Board approved the procedures and methods described below. All participants provided informed consent prior to participation and were debriefed after participation.

Shadowing Task. The shadowing task was adapted from Goldinger (1998) and consisted of three main trial types: baseline, perceptual training, and shadowing (see Figure 1). Shadowers first completed two baseline trial blocks. In baseline trials, participants were instructed to read each printed word aloud as quickly and as clearly as possible. On each trial, participants saw one of the 72 stimulus words presented on a computer monitor. Each baseline block cycled through the full 72-word stimulus list in a random order.

For both the perceptual training and shadowing phases, pre-recorded words were presented over Beyerdynamic DT-100 headphones at a comfortable listening level. Each shadower heard the full set of 72 words with each talker producing a unique set of 18 words (e.g., “bead” spoken only by SM, forming a model-word pair). Model-word pairings were counterbalanced across shadowers so that all model talker-word combinations occurred.

Perceptual training trials followed the two baseline blocks. Because shadowers were generally unfamiliar with Spanish-accented speech, shadowers were familiarized with the set of talkers in perceptual training blocks. This training block also provided listeners an opportunity to identify the most salient characteristics of each model voice to promote subsequent vocal alignment. Shadowers were exposed to both the Spanish-accented and native English-speaking model talkers before the shadowing phase began.

In the perceptual training phase, participants listened to a word spoken by one of the model talkers and then were asked to identify the word using the computer mouse in an on-screen 3x3 matrix. Although the locations of words presented in the grid and the order of words presented over the headphones were randomized, the auditorily presented word was always present in one of the nine on-screen cells. Cells of the matrix were cleared after the shadower made a selection and were not repopulated until after the next auditory item had finished playing. In each block of perceptual training trials, participants heard the list with the particular model model-word pairings that they would hear in the shadowing block. Participants heard four repetitions of the word-model talker item over the course of the perceptual training phase.

Finally, each shadower completed two blocks of shadowing trials. The full set of 72 words was represented in each block with presentation order randomized within block. Task instructions for the shadowing trials specifically excluded words like “imitate” and “repeat”, so participants were not biased towards vocal alignment by task instructions. Instead, shadowers were instructed simply to say the word they heard. Shadower utterances were recorded and edited into individual sound files for use in the subsequent AXB task described below. Recorded utterances were sampled at 22.050 kHz and amplitude normalized. Utterances from the first baseline and first shadowed block were used in the AXB task. These blocks were chosen to reduce the influence of practice effects. Utterances from the second baseline or shadowed block were used if a word was absent from the first.

The three phases of the shadowing task were administered in one testing session. All tasks were controlled via a PC computer using E-prime experiment software (Schneider, Eschman, & Zuccolotto, 2002). The entire shadowing experiment was recorded using an Audio-Technica ATR20 cardioid low-impedance microphone. Participants were tested individually in sound-attenuated rooms, and wore Beyerdynamic DT-100 headphones during the entire task (including trial blocks where no auditory stimuli were presented). After completing the shadowing task, shadowers were asked to complete a survey assessing attitudes towards immigrants and immigration. This survey was used as a gauge of attitudes toward non-native speakers of English. Because the survey responses did not appear to capture attitudes toward the particular model talkers used in the study, the survey will not be discussed further.

AXB Task. In order to obtain an overall assessment of vocal alignment, an AXB task (see Goldinger, 1998) was used. This task was first used by Goldinger (1998) to provide a global or holistic measure of alignment and has since been used in numerous studies (e.g., Miller, Sanchez, & Rosenblum, 2013; Namy et al., 2002) to index degree of alignment. This type of measure can be particularly useful because shadowers may align on a variety of acoustic dimensions that may not be captured by the measurement of individual acoustic characteristics (e.g., fundamental frequency). Sets of independent raters were presented with an AXB discrimination task in which *X* was the model talker's production of an utterance. *A* was the shadower's shadowed production, and *B* was the shadower's baseline production. Raters were asked to decide if item *A* or item *B* sounded more like item *X*. *A* and *B* were counterbalanced for order, such that a rater heard each order on half the trials. Figure 2 depicts the structure of a typical AXB task trial. If vocal alignment has occurred, raters should report at levels reliably greater than chance (.50) that the shadowed production is more similar to the model talker's production than the baseline utterance. If vocal alignment has not occurred, the shadowed and baseline productions should be judged equally similar to the model talker's production and performance should be at chance. In order to limit the number of trials in the AXB task, each rater heard utterances from a single shadower responding to all four models. A separate experiment was created for each of the 23 shadowers and 5-6 raters provided judgments for each shadower.

Results

Perceptual Learning. Accuracy on the perceptual learning task was assessed to establish whether that the model talkers for each accent group (Spanish or American English) were equally intelligible to shadowers. One challenge of using a shadowing task with accented model talkers is the increased possibility for shadowers to make errors. If shadowers successfully learned model talker characteristics during the perceptual learning block, then one might expect relatively few naming errors during the shadowing task. The perceptual learning block was first divided into four equal quartiles for assessment. Quartiles contained an equal number of items spoken by each model talker. Shadower responses to perceptual learning trials were coded as 1 for "correct" and 0 for "incorrect" and collapsed across items. This yielded a proportion correct for each shadower by model talker's accent group. Accuracy by quartile is reported in Table 1. Proportion of correct responses during the final quartile of perceptual learning trials was .99 for English talkers and .98 for Spanish talkers, indicating shadower performance was near ceiling for both talker groups. A two-way (quartile x accent) between-subjects ANOVA indicated a marginally significant effect of quartile ($F(1,154)=2.422$, $p=.0557$) and a significant effect of accent ($F(1,154)= 7.225$, $p=.008$). The interaction of quartile and accent was not significant ($F(1, 154)= 1.133$, $p=.338$). The significant main effect of accent indicates that accuracy was significantly lower for the Spanish talkers. Lack of reliable difference between quartiles suggests that performance did not improve over the course of the training.

Shadowing Task Accuracy. Although performance on the perceptual learning task was high, the shadowing task has different demands (i.e. searching a grid for a printed word versus naming the word verbally). As an additional check as to whether or not the English and Spanish model talkers differed in intelligibility, accuracy of shadowing responses was evaluated. To determine shadowing response accuracy, one of five trained research assistants judged whether or not a token produced by a shadower during the first shadowing block was an accurate production of that word category. If a rater judged the token to be an error, she marked it as such. Recall, however, that two blocks of shadowing responses were collected. If an error was made in the first shadowing block, the researcher searched for an acceptable token in the second shadowing block. If the shadower made a mistake in both the first and second shadowing block, the item was excluded from analysis and from the AXB task for that shadower. For the present analysis, items that a shadower missed in both the first and second shadowing blocks are counted as errors. By this metric, 6.8% ($^{112}/_{1,656}$) of shadowing trials resulted in errors. Error rates by model talker is reported in Figure 2. Of the trials that were errors, shadowers made proportionately fewer errors when the model was a native English speaker than when the model was a native Spanish speaker ($\bar{x}_{\text{English}}$=12.5%; $\bar{x}_{\text{Spanish}}$=87.5%). While the overall incidence of naming errors was low, accent group appears to be systematically related to shadower error rates ($X^2(1)=63$, $p<.001$).

Shadowing Task Vocal Alignment. Vocal alignment was assessed using two types of measures: acoustic and listener judgment. For both types of analyses, the main effect of interest is the influence of model talker accent on perceived vocal alignment.

Acoustic Assessment. Two acoustic measures of vocal alignment were collected, fundamental frequency (f0) and utterance duration (duration). These measures were used to establish whether shadowers vocally aligned to the model talkers along acoustic dimensions indexing changes in pitch and speaking rate. F0 and duration were computed using Praat (Boersma & Weenink, 2014) and measures for baseline and shadowed utterances were compared. To index alignment, difference scores were calculated by subtracting the baseline value from the shadowed value (e.g. shadowed f0 – baseline f0) for each word token. When calculated in this way, the sign (positive or negative) indicates whether the shadowed token had larger or smaller magnitude than baseline token. For example, a difference score of 4.5 Hz would reflect a shadowed f0 greater than baseline f0. Magnitude of the difference score expresses extent of change, with larger values illustrating more change than smaller values. Mean f0 difference scores and mean duration difference scores for each model talker group are reported in Figures 3A and 3B, respectively. The data in Fig. 3A indicate a trend towards lower f0 values from baseline to shadowed utterance across both accent types. Fig. 3B shows that shadowers increased their duration from baseline to shadowed utterance for both accent types.

To determine if talker accent had an effect on the acoustic measures of vocal alignment, a mixed-effects model was created for each dependent measure. Mixed-effects modeling (MEM) confers many advantages over traditional analysis methods, particularly because of its ability to incorporate multiple fixed (i.e., independent variables) and random (i.e., participant or item) effects (Baayen, Davidson, & Bates, 2008; Breslow & Clayton, 1993; Jaeger, 2008). Because of this flexibility, it is possible to account for participant and item effects in a single analysis. The MEMs reported here were conducted in the `R` environment (R development core team, 2013) using the `lme4` (Bates, Maechler, Bolker, & Walker, 2014) and `languageR` (Baayen, 2011) packages. Control MEMs for all tests included random effect terms only. Significance of each independent variable was determined through conventional *t*- and *z*-statistics, and through step-wise model comparison (Chi-square test) of MEMs where the first comparison was to the control MEM.

The MEM for f0 difference score will be described first. In order to determine if the model talker's f0 influenced vocal alignment as indexed by shadowers' f0 difference scores, models including and excluding model talkers' f0 were constructed. The control MEM included random effects only (shadower and word), as described above. The intercept of the control MEM was non-significant (β=-4.292; $t(23.809)$=-1.113, p=.277), indicating that the mean f0 difference score was not significantly less than 0. Since f0 difference score did not differ from 0, no further analyses were conducted on f0.

A second acoustic metric of vocal alignment was duration difference score. A control MEM with duration difference score as the dependent measure and random effects was constructed. Duration difference score was significantly greater than 0, as indicated by the positive intercept in the control MEM (β= 34.13; $t(23)= 3.303$, p=.003). Accent was added into a test model to assess its contribution to duration difference score. The effect of accent was significant (β= -9.537; $t(1527.3)= -2.165$, p=.031), and the sign of the intercept indicates that the difference score was greater for English talkers than for Spanish talkers. Comparison to the control MEM indicated that accent significantly improved the fit of the model ($X^2(1)= 4.68$, $p<.031$). One plausible concern is that shadowers were not sensitive to the talker's duration specifically, but that this effect is driven by a general tendency of the shadowers to slow down and lengthen their utterances over the course of the experiment. To test if this was the case, model talker's duration was added into the test model to examine whether duration difference score was related to real variation present in model talker's speaking rates. Longer talker durations were significantly related to bigger duration difference scores (β=.115; $t(1527.5)=7.016$, p<.001). Adding model talker's duration along with accent improved the MEM's fit ($X^2(1)= 48.67$, $p<.001$. These findings suggest that shadowers were sensitive to model talker's utterance duration and that alignment varied across accent type. More broadly, the acoustic analyses indicate that duration difference score, but not f0 difference score, reflected vocal alignment.

AXB Assessment. Although individual acoustic measures of vocal alignment can be informative, alignment as assessed by human raters can capture the manifold variations in speech that single acoustic dimensions cannot (see Goldinger, 1998).

Therefore, vocal alignment as measured by AXB listener judgment was tested as well. Responses for each trial of the AXB task were coded as 0 for a *BX* judgment (baseline is more similar to model) and 1 for an *AX* judgment (shadowed is more similar to model) prior to analysis. Since a judgment that the shadowed item is closer to the model's item is represented by this coding scheme, it results in a proportion of perceived vocal alignment. Mean AXB accuracy is reported in Figure 4.

To determine whether AXB accuracy was significantly greater than chance, a control logistic MEM (a subtype of MEM for binary outcomes) was calculated. In addition to the random effects of shadower and word, the nature of the AXB task introduces listener as random effect. The overall likelihood that a listener chose a shadowed item as being more similar to a model token than to a baseline token was .55 (sd=0.058), which the intercept in the control model confirmed was significantly greater than chance (β= .207, Z = 3.823 p<.005). Model talker's accent was then added to a test MEM. Accent had a significant effect on AXB judgment (β_{accent}= .11, z = 4.899 p<.001). Because English accent was used as the reference category for talker's accent, the β estimate indicates that there was greater alignment towards Spanish-accented model talkers than to American-accented talkers. Comparison against the control MEM indicated that talker's accent improved the fit to the data ($X^2(1)$= 24.015, p<.001).

While the primary aim of the experiment was to assess the effect of accent on perceived vocal alignment, it is necessary to consider how acoustic cues relate to global perceptions of alignment. As such, a second logistic MEM was created with AXB judgment as the dependent measure. In addition to accent, duration difference score and f0 difference score were added step-wise into the MEM as fixed effects. Both duration and f0 difference scores were *z*-transformed (i.e., centered) before being entered into the MEM. The specified random effects were shadower, word, and listener. β-estimates, *Z*-statistics, and *p*-values associated with the second MEM are presented in Table 2. The significant intercept indicates that AXB accuracy was above chance. Duration difference score and accent are significantly related to AXB accuracy, but f0 difference score is not. Model testing confirmed that adding centered duration difference score improved the fit of the MEM, but adding centered F0 difference score had no effect. When compared to the difference scores MEM, the effect of accent in the AXB MEM is reversed, such that shadowers aligned more to Spanish-accented talkers than to native-accented talkers. The interaction between F0 difference score and duration difference score was also non-significant ($X^2(1)= 1.223$, $p=.269$), and was excluded from the MEM. This MEM indicates that duration and model talker's accent, but not F0, contribute to the likelihood that vocal alignment was perceived.

Discussion

The present study sought to characterize the processes underlying vocal alignment behavior using model talker's accent to investigate social influences on alignment. Accented speech carries social information during spoken communication and has relevance to native speakers of a language. Under a social theory of alignment, the social information acting on accent could have led to less vocal alignment to accented talkers, but from a perceptual-motor perspective, the difference between native and non-native speech production norms should result in more alignment. To collect measures of vocal alignment, participants shadowed four highly intelligible model talkers from one of two accent groups, Standared American- and Spanish-accented. Alignment was assessed separately using acoustic analyses and listener judgment, and then coupled by using acoustic dimensions as additional predictors of listener judgment. With respect to acoustic measures of alignment, fundamental frequency did not undergo reliable change, contrary to work that indicates an important role of f0 in vocal alignment (e.g., Babel & Bulatov, 2011). Speaking rate, as measured by word duration, was a reliable measure of alignment, similar to work that has found alignment with systematic manipulation of duration (Staum-Casasanto et al., 2010). Listeners in the AXB task were also able to detect alignment. The main hypothesis of the experiment, that native English shadowers would demonstrate different degrees of alignment to non-native accent groups and native accent groups, was supported.

However, the effect of accent varied depending on the type of measure used to assess alignment. An acoustic measurement of vocal alignment indicated that shadowers aligned more to native than non-native model talkers. When measured by AXB judgment, the trend was reversed, and more alignment was detected toward non-native models than native models. The disagreement between the acoustic analysis and AXB analysis reinforces that listener judgments are made based on a multidimensional suite of acoustic features, rather than simply one or two dimensions. This highlights the importance of collecting listener ratings in addition to taking acoustic measurements (Pardo, 2013; Pardo et al., 2013). Importantly, the finding indicates that having a noticeable non-native accent does not prevent native talkers from aligning.

From a social-motivational perspective, the results might imply that attitudes toward high intelligibility talkers of a non-native accent group are positive or relatively neutral, despite previous work suggesting negative attitudes toward accented speakers in general and Spanish accented English in particular (McKirnan & Hamayan, 1984; Ryan, 1983). Other work however suggests that attitudes towards talkers tend to be more positive if the talker is more intelligible (Hanson et al., 2004; Ryan et al., 1977). If a foreign accent does not substantially lower a talker's intelligibility, then negative attitudes on the part of a shadower might not be engaged, thus prompting native-like patterns of alignment. This could explain the disparate findings on alignment to non-native accented talkers (Kim, 2011; Van Engen et al., 2010). A neutral or positive view of the talkers could also have been the result of the perceptual learning phase. If familiarity with a talker can reduce the impact of stereotype towards a talker group, then perhaps the presence of perceptual learning influenced shadower attitudes. The perceptual learning task may not have been demanding enough to completely eliminate shadowing errors, but could have been effective in acquainting the shadowers with each talker, thus promoting equivalent degrees of alignment. However, the present study did not observe equal rates of alignment to native and non-native accented talkers. When assessed via listener judgment, vocal alignment was greater for Spanish-accented models. This pattern is more consistent with a perceptual-motor perspective of vocal alignment than a social-motivational one.

As described earlier, a prediction made by a perceptual-motor account is that vocal alignment will be greater when distance between the target and the shadower's baseline is greater (Fowler & Galantucci, 2005; Kim et al., 2011, Shockley et al., 2004). Accentedness ratings of the talkers used in the current study suggest that the Spanish models were perceived as having Spanish accents. It is possible that alignment toward the

Spanish-accented speech was due to the mismatch between the shadower's acoustic linguistic representations and the model's. Because the Spanish models were noticeably accented, their speech production was likely different from an American English norm. The greater distance between the Spanish-accented and native-accented norms could have resulted in greater alignment toward the Spanish models. Under circumstances that involve minimal social context, then, a speech perception-speech production link could explain the differential rates of alignment to native and non-native speech.

Nevertheless, the possibility that attitudes towards talker group were not invoked leaves open the interpretation that the present results were the effect of neutral opinions toward all models. The findings could be more completely explained from a general behavioral mimicry perspective. According to a behavioral mimicry view, the automatic priming of social-motivational information causes all types of behavioral alignment, and a cognitive control mechanism can inhibit alignment when needed (Van Baaren et al., 2009). The implication for vocal alignment is that alignment will occur unless specific constructs are primed to prevent alignment or to promote divergence. For the present study, high intelligibility talkers of two accent groups served as models. If, as suggested above, the intelligibility or degree of accentedness did not prime negative attitudes toward any of the talkers, then the system default would still elicit vocal alignment. To explain why listeners judged more alignment to Spanish-accented talkers, the same reasoning from the motor account could be invoked; because there was more 'distance' between the model and shadower at the outset, more change was observed. The difference is that the behavioral mimicry account does not assume that the retrieval of speech gestures is necessary for alignment. Instead, distance along any behavioral metric of alignment would evoke vocal alignment in the absence of a social-motivational directive not to align. Because the behavioral mimicry account combines aspects of the social-motivational and perceptual-motor views, it constitutes a more integrated explanation of the current findings.

Given that the findings from the present study are based on highly intelligible talkers, caution should be taken when generalizing the results. The varied patterns of alignment to foreign-accented talkers found in prior work suggest that talker intelligibility and accentedness may matter (Bent et al., 2006; Kim et al., 2011; Van Engen et al., 2010). The number of model talkers, while consistent with the number typically examined in shadowing studies, captures only a small amount of possible model talker variation. With only one Spanish male, one Spanish female, one English male, and one English female, it is difficult to claim that alignment patterns will be the same for all other English and other Spanish talkers. Systematically controlling model talker intelligibility and adding more model talkers could increase generalizability of vocal alignment patterns. It is also important to consider the variation that shadowers bring to the experimental setting. For example, evidence suggests that males and females align to different extents and to different dimensions (Bilious & Krauss, 1988; Namy et al., 2002; Pardo, 2006). The majority of shadowers in the present study were female, and not enough males participated to allow for comparison between sexes. Future studies whose primary aims are to examine sex differences in alignment behavior should consider a more balanced shadower sample. Although sex differences could potentially account for some variance, individual differences are an important avenue for future research.

One commonality among shadowing studies is that shadowers differ considerably in their alignment behavior. The specific factors that contribute to this individual variability are important topics for future research. For example, extraversion may correlate with alignment behavior. Extraverts are typically sociable and outgoing (Goldberg, 1992; Tupes, & Christal, 1961). A shadower high on extraversion may be more sensitive to the speech of others. If extraversion is positively correlated with attention to speech, then one might expect that extraverted shadowers would be more likely to exhibit vocal alignment than introverted shadowers. Effects of extraversion specifically and other individual characteristics generally have yet to be tested.

Individuals' attitudes towards particular talkers or groups of talkers could also be examined using methods other than self-report. Because attitudes towards stigmatized speaker groups tends to be a sensitive topic, the present study's attitude measure may have been inappropriate or ineffective for appraising shadowers' true opinions. One measure that has been used with some success in assessing implicit attitudes is the Implicit Association Test (IAT) (Babel, 2010; Greenwald, McGhee, & Schwartz, 1998). Following work by Babel (2010), one could additionally manipulate attitude towards a specific accented talker through the use of contrived scenarios. Tests examining and modulating attitudes would provide useful evidence for social-motivational accounts of alignment. If the effects were instead examined in a shadowing paradigm, which would minimize other sources of social information, a claim could be made in support of implicit social priming.

Though there are still avenues for further exploration, the present study finds support for a view of vocal alignment that involves both social-motivational and automatic components. The combination of these factors results in a change in speech production towards both accented and non-accented model talkers even when social cues are minimized.

References

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological science*, *21*(12), 1903-1909.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177-189.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society, 39*, 437-456. DOI: 10.1017/S0047404510000400

Babel, M., & Bulatov, D. (2011). The role of fundamental frequency in phonetic accommodation. *Language and Speech, 55*(2), 231-248.

Ball, P., Giles, H., Byrne, J. L., & Berechree, P. (1984). Situational constraints on the evaluative significance of speech accommodation: Some Australian data. *International Journal of the Sociology of Language*, *1984*(46), 115-130.

Barr, D.J, Levy, R., Scheepers, C., Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255-278.

Bent, T., Bradlow, A.R., & Smith, (2006). Phonetic convergence to accented speech.

Baayen, R. H. (2011). languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics". R package version 1.4. (http://CRAN.R-project.org/package=languageR).

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Bates, D., Maechler, M, Bolker, B. & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-4. (https://github.com/lme4/lme4/)

Bilous, F.R. & Krauss, R.M. (1988). Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. *Language and Communication, 8(*3/4), 183-194.

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.3.68, retrieved 20 March 2014 from http://www.praat.org/

Bourhis, R. Y., & Giles, H. (1977). The language of intergroup distinctiveness. *Language, ethnicity and intergroup relations*, *13*, 119.

Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Society, 88(*421), 9-25.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707-729.

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America*, *106*, 2074.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic coordination in dialogue. *Cognition, 75*, B13–B25.

Branigan, H.P., Pickering, M. J., Pearson, J., McLean, J.F., Brown, A. (2011). The role of beliefs in lexical alignment: evidence from dialogs with humans and computers. *Cognition, 121,* 41-57.

Brouwer, S., Mitterer, H., & Huettig, F. (2010). Shadowing reduced speech and alignment. *Journal of The Acoustical Society of America, 128*, EL32-EL37.

Buchan, N. R., Johnson, E. J., & Croson, R. T. (2006). Let's get personal: An international examination of the influence of communication, culture and social distance on other regarding preferences. *Journal of Economic Behavior & Organization*, *60*(3), 373-398.

Chartrand, T. L. & Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6) 893-910.

Doddington, G. R. (1985). Speaker recognition—Identifying people by their voices. *Proceedings of the IEEE*, *73*(11), 1651-1664.

Evans, B.G. & Iverson, P. (2007). Plasticity in vowel perception and production: a study of accent change in young adults. *Journal of the Acoustical Society of America, 121*(6), 3814-3826.

Fowler, C. A., & Galantucci, B. (2005). The relation of speech perception and speech production. *The handbook of speech perception*, 632-652.

Giles, H. (1973). Accent mobility: a model and some data. *Anthropological Linguistics, 15(*2), 87-105.

Giles, H., Coupland, N., & Coupland, J. (1991). *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge: Cambridge University Press.

Giles, H., & Ogay, T. 2007. Communication accommodation theory. In Bryan B. Whaley & Wendy Samter (eds.), *Explaining Communication: Contemporary Theories and Exemplars*, 293– 310. Mahwah, NJ: Lawrence Erlbaum.

Gluszek, A., & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Pers. Soc. Psychol. Rev.*, *14*(2), 214-237.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological assessment*, *4*(1), 26.

Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251-279.

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 152.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464-1480.

Gregory, D., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status predictions.

Journal of Personality and Social Psychology, 70, 1231–1240.

Hanson, E. K., Beukelman, D. R., Fager, S., & Ullman, C. (2004). Listener attitudes toward speech supplementation strategies used by speakers with dysarthria. *Journal of Medical Speech-Language Pathology*, *12*(4), 161-166.

IEEE Subcommittee (1969). “IEEE recommended practice for speech quality measurements,” *IEEE Trans. Audio Electroacoustics, 17*, 225–246.

Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory & Language, 59*, 434-446.

Kappes, J., Baumgärtner, A., Peschke, C., and Ziegler, W. (2009). Unintended imitation in non-word repetition. *Brain and Language, 111*, 140-151. DOI: 10.1016/j.bandl.2009.08.008

Kim, M. (August, 2011). *Phonetic convergence after perceptual exposure to native and nonnative speech: preliminary findings based on fine-grained acoustic-phonetic measurement.* Unpublished paper presented at The 17th International Congress of Phonetic Sciences, Hong Kong, China.

Kim, M., Horton, W.S., & Bradlow, A.R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology, 2(*1), 125–156. DOI: 10.1515/labphon.2011.004

Kučera H., and Francis W. N. (1967). Computational Analysis of Present-Day American English (Brown University Press, Providence, RI).

Lane, H. (1963). Foreign accent and speech distortion. *Psychology Faculty Publications*.,Paper 10. http://hdl.handle.net/2047/d20000857

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review*, *74*(6), 431.

Liberman, A.M. & Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Louwerse, M. M., Dale, R., Bard, E.G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, 1-24.

Luce P. A., and Pisoni D. D. (1998). "Recognizing spoke words. The neighborhood activation model," *Ear and Hearing*, *19,* 1–36.

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 676– 684.

McKirnan, D. J., & Hamayan, E. V. (1984). Speech norms and attitudes toward outgroup members: A test of a model in a bicultural context. *Journal of Language and Social Psychology*, *3*(1), 21-38.

Miller, R. M., Sanchez, K., & Rosenblum, L. D. (2013). Is speech alignment to talkers or tasks?. *Attention, Perception, & Psychophysics*, *75*(8), 1817-1826.

Mitterer, H., & Ernestus, M. (2008) The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition, 109*, 168-173.

Mitterer, H., & Müsseler, J. (2013). Regional accent variation in the shadowing task: Evidence for a loose perception-action coupling in speech. *Attention, Perception, & Psychophysics, 75*, 557-575.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73-97.

Namy, L.L., Nygaard, L.C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: the role of perception. *Journal of Language and Social Psychology, 21(*4), 422-432.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*(3), 355-376.

Nygaard, L.C., & Queen, J.S. (2000). The role of sentential prosody in learning voices. *The Journal of the Acoustical Society of America, 107*(5), 2856.

Pardo, J.S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119(*4), 2382-2393.

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics*, *72*(8), 2254-2264.

Pardo, J. (2013, June). Reconciling diverse findings in studies of phonetic convergence. In *Proceedings of Meetings on Acoustics* (Vol 19, p. 060140).

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2011). Phonetic convergence in college roommates. *Journal of Phonetics, 40*, 190-197.

Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: the relation between acoustic and perceptual measures. *Journal of Memory and Language*, *69*(3), 183-195.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America, 24,* 175–184.

Pickering, M.J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27,* 169-226.

R development core team (2013). R: A language and environ- ment for statistical computing. Vienna: R Foundation for Statistical Computing, http://www.R-project.org.

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1019.

Rubin, D. L., & Smith, K. A. (1990). Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of nonnative English-speaking teaching assistants. *International Journal of Intercultural Relations*, *14*(3), 337-353.

Ryan, E.B. (1983). Social psychological mechanisms underlying native speaker evaluations of non-native speech. *Studies in Second Language Acquisition, 5,* 148-159. doi:10.1017/S0272263100004824.

Ryan, E. B., Carranza, M. A., & Moffie, R. W. (1977). Reactions toward varying degrees of accentedness in the speech of Spanish-English bilinguals. *Language and Speech*, *20*(3), 267-273.

Sancier, M.L. & Fowler, C.A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Jounal of Phonetics, 25*, 421-436.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Psychology Software Incorporated.

Sidaras, 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.

Sheffert, S. M., Pisoni, D. B., Fellowes, J. M., & Remez, R. E. (2002). *Journal of Experimental Psychology: Human Perception and Performance*, *28*(6), 1447.

Shockley, K., Baker, A.A., Richardson, M.J., & Fowler, C.A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception and Performance, 33, (*1), 201–208. DOI: 10.1037/0096-1523.33.1.201

Shockley, K., Sabadini, L., & Fowler, C.A. (2004). Imitation in shadowing words. *Perception & Psychophysics 2004, 66 (3), 422-429.*

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass*, *3*(1), 236-264.

Staum-Casasanto, L., Jasmin, K., & Casasanto, D. (2010). Virtually Accommodating: Speech rate accommodation to a virtual interlocutor. In S. Ohlsson & R. Catrambone (Eds.), Proceedings of the 32nd Annual Conference of the Cognitive Science Society (pp. 127-132). Austin, TX: Cognitive Science Society.

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition*, *119*(1), 131-136.

Tupes, E. C., & Christal, R. E. (1961). *Recurrent personality factors based on trait* *ratings*. Technical Report ASD-TR-61-97. Lackland Air Force Base, TX: U.S. Air Force.

Uther, M., Knoll, M.A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication, 49,* 2-7.

Van Baaren, R., Janssen, L., Chartrand, T.L. & Dijksterhuis, A. (2009). Where is the love? The social aspects of mimicry. *Philosophical Transactions of the Royal Society B, 364*, 2381-2389.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, *53*(4), 510-540.

Table 1

Proportion of correct responses on the perceptual training task by quartile.

		Quartile			
		1	2	3	4
Accent	English	0.9892	0.9892	0.9929	0.9899
	Spanish	0.9762	0.9870	0.9916	0.9786

Table 2

MEM Estimates Associated with the MEM using Accent, Duration Difference Score, and f0 Difference Score to Predict AXB Judgment.

Fixed Effects	β	*se*	*Z*	*p*(*Z*)	X^2	*p*(X^2)
(Intercept)	.155	.051	3.065	.002		
Accent [Spanish]	.119	.023	5.252	.000	24.015	.000
zDurationDS	.123	.013	9.232	.000	265.55	.000
zf0DS	.113	.014	-0.917	.359	.8402	.359

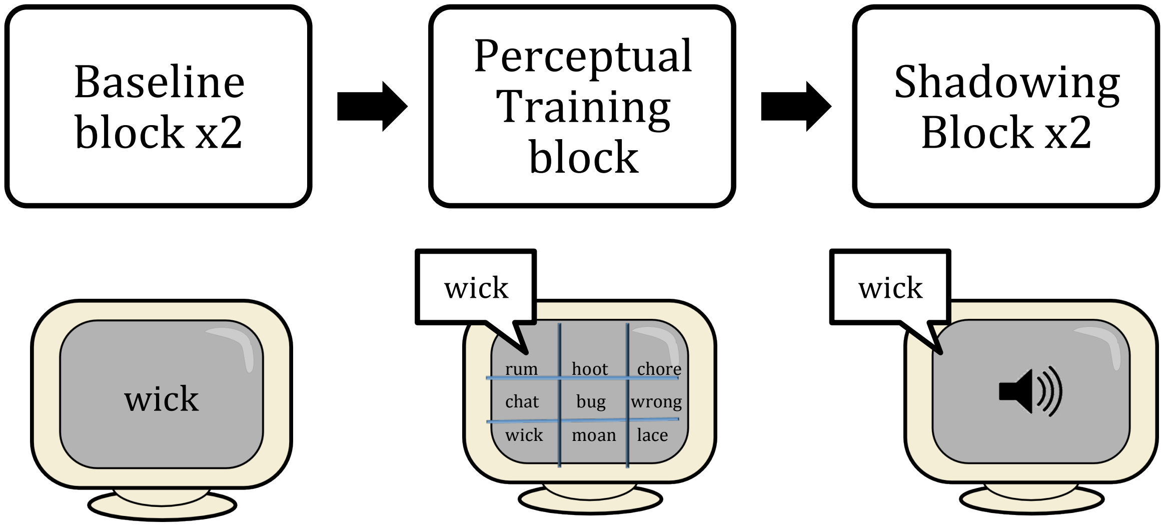

Figure 1. The top row is a schematic of the shadowing task's construction. The bottom row depicts what a participant might see on a given trial within the corresponding block type.

Figure 2. Rates of errors made for each model talker during shadowing trials, with standard error bars.

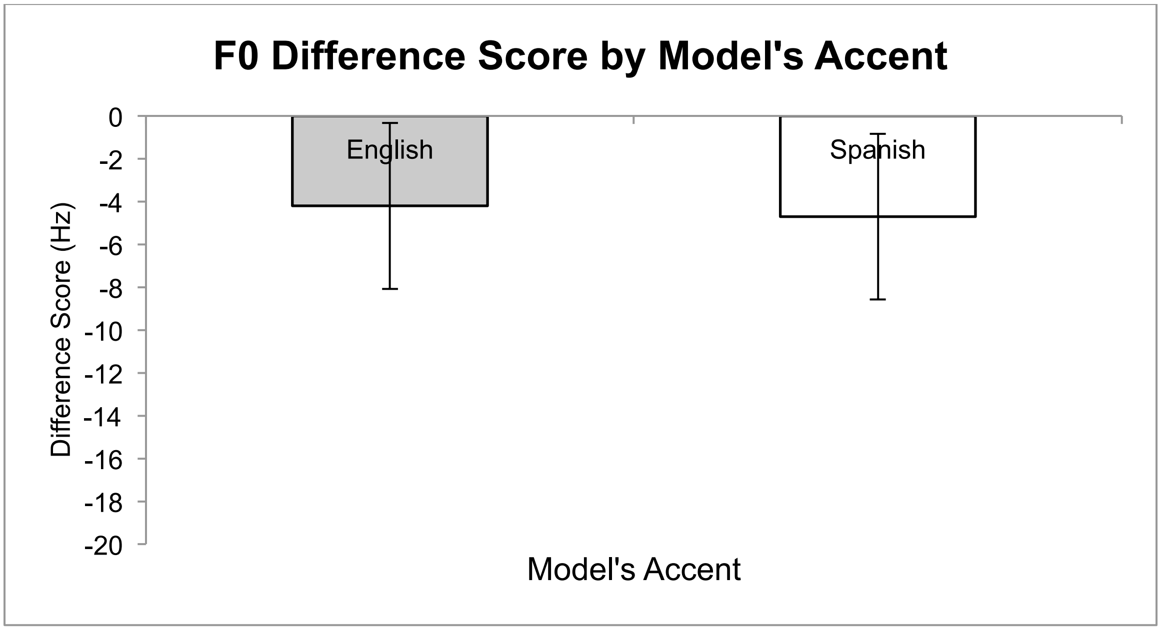

Figure 3A. F0 difference score by model talker’s accent, with error bars representing standard error of the mean.

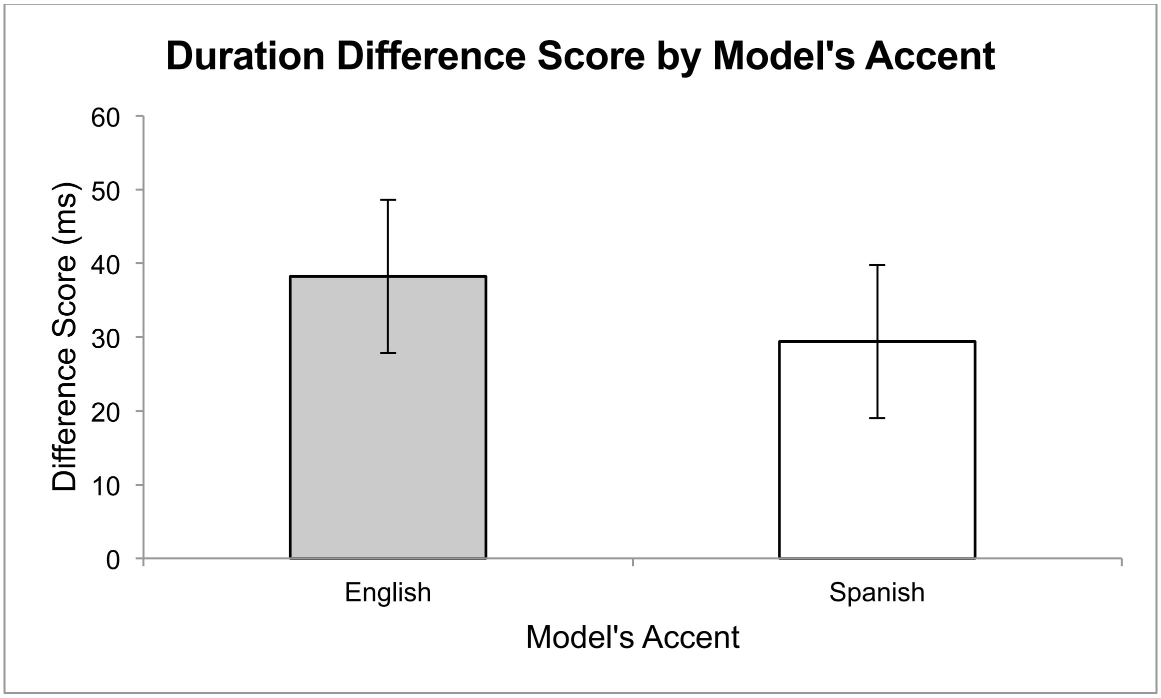

Figure 3B. Duration difference score by model talker's accent. Error bars represent standard error of the mean.

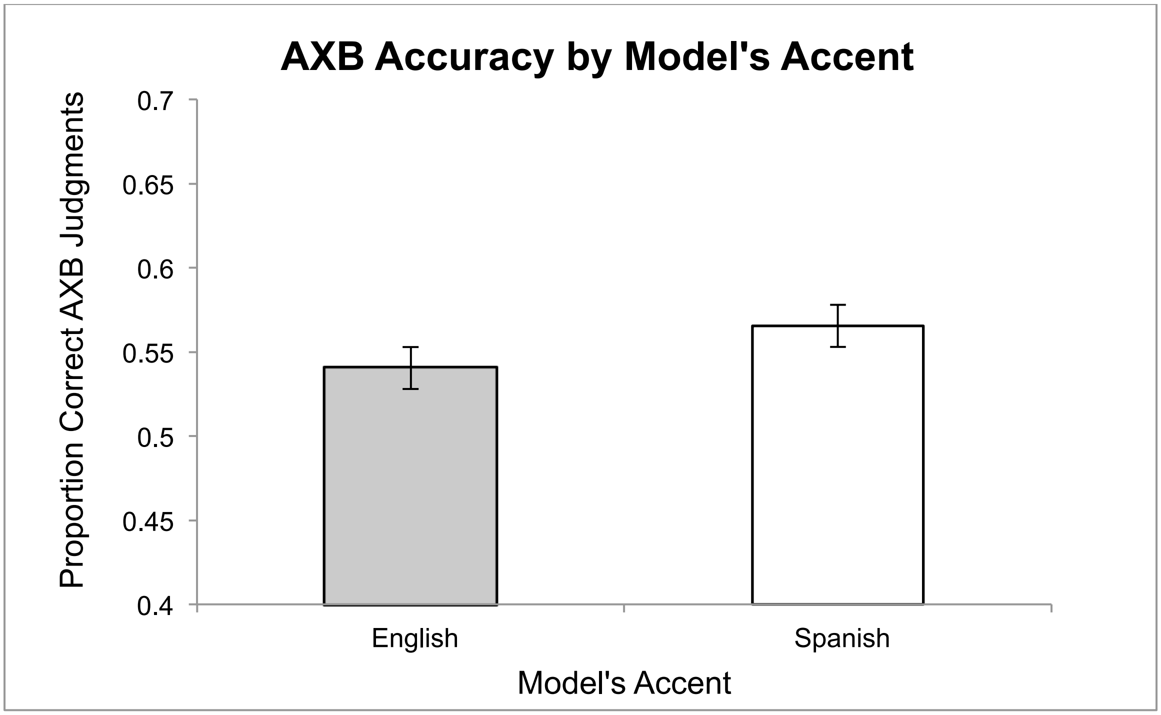

Figure 4. Accuracy of AXB judgments by model talker's accent. Here, a *correct* judgment is rating a shadowed item as more similar to the model, rather than selecting a baseline item as more similar to the model. Error bars represent standard error of the mean.

Appendix: Full Word List

List of *hard* words used in both visual and auditory presentation formats.

ban	dune	lice	rim
bead	fade	mace	rum
beak	fin	main	rut
bean	goat	mall	sane
bud	gut	mat	sill
bug	hack	mid	soak
bum	hag	mitt	tan
bun	hash	moan	teat
chat	hick	moat	toot
cheer	hid	mole	wad
chore	hoot	mum	wade
cod	hum	pad	wail
comb	kin	pat	wed
con	kit	pawn	weed
cot	knob	pet	white
dame	lace	pup	whore
den	lad	rat	wick
doom	lame	rhyme	wrong