THE EFFECT OF FUNDAMENTAL FREQUENCY ON PHONETIC CONVERGENCE

Dasha Bulatov
Department of Linguistics
University of California, Berkeley

ABSTRACT

This paper examines the importance of fundamental frequency (F0) in the process of phonetic convergence within an immediate-repetition or "shadowing" task. Previous research has suggested that F0 facilitates the transmission of social information that individuals can use to establish their social orientation within an interaction (Gregory et al., 1991, 1996, 2001). Social theories of accommodation assert that this process mediates a subconscious decision to converge (imitate) or diverge in speech, and to what extent. If this is true, having participants shadow a talker whose speech is high-pass filtered should demonstrate that they imitate less than participants who shadowed the talker's full range of speech. Additionally, this project will review social and automatic theories of convergence, and reinforce the need for a new, integrative model that allows social processes to intervene in an exemplar-based perception-production link.

INTRODUCTION

The theoretical framework for this research is primarily situated in two related models that explain accommodation as stemming from social behavioral motivations. The first, Communication Accommodation Theory (CAT), proposed by Giles and Coupland (1991), interprets accommodation as an indication that interlocutors adapt to each other's communicative behavior as a strategy achieve a desired social distance. This model emphasizes the social function of language within a larger set of communicative behaviors motivated by an individual's desire to identify with or seek acceptance in a particular social group. The second model, The Vocal Channel Social Status Model (VOCSTAT) was developed by Gregory et al. (2001) in response to their research that suggested the importance of fundamental frequency in communicating the social information that individuals would accommodate to within the CAT framework. It posits

that the vocal channel, specifically F0, is a crucial pathway for conveying that information.

BACKGROUND

I. *Communication accommodation theory (CAT)*

CAT, described by Giles and Coupland (1991), frames language as a tool that demonstrates and aligns social relations through measures of social approval. It is in line with social identity theory (Tajfel and Turner, 1979) insofar as individuals define their identity by categorizing others and themselves and others into groups that are constantly being compared and evaluated positively or negatively. The group to which an individual belongs is evaluated more favorably than others. CAT specifies how linguistic behavior can contribute to the maintenance of such identities via modifying speech to sound more or less like one's conversational partner.

In this theory, it is important to distinguish between *convergence* and *accommodation*. Convergence is "a strategy whereby individuals adapt to each other's communicative behaviors in terms of a wide range of linguistic/prosodic/non-verbal features," (Giles and Coupland, 1991: 63) while accommodation is "the general sense of adjusting our communication actions relative to those of our conversation partners" (p. 60). Therefore, *convergence* emphasizes the merger of vocal and phonetic elements (such as intonation contours and vowel space) while *accommodation* emphasizes the performance of a service by one conversational partner for another.

The ability of talkers to shift their production towards (or away from) the productions of their conversational partner is a crucial way in which speakers can situate themselves within a conversational space, which in turn represents a larger, more inclusive social space. Talkers can modify their patterns of production to varying degrees as a function of (i). social factors that characterize individuals or groups and (ii). situational and social contexts in which an utterance is produced. In this framework, phonetic convergence can be considered part of a larger set of social-behavioral strategies in which individuals sub-consciously engage to achieve particular social goals, such as approval or inclusion. Language is one tool that individuals can manipulate in order to position themselves in a social context. Convergence, then, results when social factors

align in such a way that the accommodative instinct of the individual is to alter his speech production to resemble that of his conversational partner, presumably because he or she shares an identity with that partner and wants to decrease the social distance between them (Giles and Coupland, 1991). Convergence interacts with the social variables of status and dominance and the accommodative interplay that results between.

Additionally, CAT describes three other accommodation strategies that interlocutors can adopt during the course of a conversation: maintenance, complementarity and divergence (Giles and Coupland, 1991). Divergence occurs when linguistic differences are increased by speakers in order to create social separation. Maintenance involves no modification or shift in production, and complementarity describes a situation in which two speakers emphasize social distance between themselves using their language. In the immediate-repetition paradigm that will be employed in this study (where one talker repeats after another immediately after hearing his or her production), phonetic analysis performed on the speech of the participant can reveal evidence for these types of behaviors, and is a useful way to track variability in minute detail, such as voice-onset time or formant values. Using listener judgments is a more holistic method to reveal accommodative behavior, although a focus on subjective human interpretations of accommodation can introduce a variety of complicating factors, as will be seen in the analysis of the data collected for this study.

Bourhis and Giles have led research in support of CAT in studies of accent convergence and divergence. A study was conducted on two groups of Welsh-born adults who were both enrolled in Welsh language classes, but for different purposes (Bourhis and Giles, 1977). The first group -the "integrative" learners- wanted to learn the Welsh language to explore and preserve their cultural heritage, and the second group -the "instrumental" learners- was interested in learning it only to augment their business and career opportunities. Individuals in both groups were interviewed by a Received Pronunciation-accented male who expressed doubts about the vitality of the Welsh language. The "integrative" learners responded to the identity-threatening questions by accentuating linguistic differences between themselves and the interviewer, primarily by increasing how "Welsh" their accent sounded. On the other hand, the "instrumental" learners converged towards the RP interviewer, suggesting that they did not value their

Welsh identity as highly. Two linguistically naïve and untrained judges determined the perceived changes in accent for this experiment using eleven criteria.

In another study conducted in a French-Canadian community, bilingual speakers who attempted to accommodate to English-dominant listeners by speaking English were rated more favorably, suggesting that accommodation increases the quality of an interaction (Giles et al., 1993). Giles also demonstrated the saliency of accommodation in interactions by investigating whether interviewees would converge their speech towards an interviewer for social approval (Giles 1973). He did so by conducting interviews with 13 boys using himself (an RP speaker) and a 17 year-old boy as the interviewers. Accent changes were judged by 18 Welsh-born adults and 18-Bristol born adults. Giles found that interviewees converged their speech towards the interviewer in all circumstances, suggesting that the desire for social approval is a salient motivation for convergence towards an authoritative figure.

Other studies that reveal the role of language as an identity marker and linguistic accommodation as a means of reinforcing that identity include a survey of Puerto-Rican adolescents who described language as a salient marker of ethnic identity (Giles et al., 1979) and a study of Chinese-English bilinguals who, on a questionnaire, identified more with Chinese values when the questions were posed in Chinese rather than English (Yang and Bond, 1980). Overall, the above-mentioned studies, as well as related research, provide a great body of evidence that supports CAT's assertion that linguistic modification through accommodative strategies depends on social variables such as perceived identity.

In addition, CAT offers a method for determining power and status differentials that may exist between interacting persons or within the social context of an interaction. It assumes that accommodation will occur over time, once the status relations of the interlocutors have been ascertained. Additionally, any status differences should lead to an asymmetry in accommodation; a person with lower status should accommodate the person of the higher status. Gregory and Webster (1996) tested the prediction that verbal accommodation should be moderated by relative social status by analyzing interviews between television host Larry King and guests with varying degrees of perceived social status. If the guest has a lower perceived social status than King, he or she will

accommodate to him. On the other hand, if King interviews a guest with very high status, *he* should accommodate to him or her. The lower frequency band of each interview partners' speech was analyzed for convergence, and indeed their findings demonstrated that asymmetry of convergence corresponded to a status differential between partners. That is, King accommodated to high-status guests much more than they accommodated to him, and low-status guests tended to accommodate to King. Their results support the CAT characterization of convergence as a way for a speaker to obtain a sense of social inclusion as well as its prediction that "the greater the speaker's need to gain another's social approval, the greater the degree of convergence will be" (1991: 73).

Accommodation also contributes to the perceived quality of a conversation. As discussed in Gregory and Webster (1996), acknowledgment of power differentials through appropriate levels of accommodation (where a lower status person converges to a higher status person, for example) is crucial in maintaining an effective, or optimal, communication environment. This assumption was supported by a series of studies in which the quality of interviews was rated according to 34 different indices (Gregory et al., 1991, 1993, 1997, see below). Interviews were rated more favorably by listeners if they perceived that convergence had occurred.

II. *Vocal Channel Social Status Model (VOCSTAT) and the role of F0*

First outlined in 2001 by Gregory and his colleagues, this model represents a theoretical merger between the Communication Accommodation Theory and Patterson's Sequential-Functional Model (SFM) of communication (Patterson, 1983). The latter model is a comprehensive attempt to integrate verbal and nonverbal expressions with the effects of "antecedent factors" (such as age, gender and background) brought to the interaction by the participants and "preindication mediators" (such as behavioral predispositions and cognitive-affective assessments) that influence expectations about the current interaction (Burgoon et al., 1995). The "interaction phase" is influenced sequentially and often subconsciously by antecedent factors and preindication mediators. The latter can be revised over the course of the interaction and includes relevant social variables such as dominance and status, and results in accommodation. These assessments, however, increase cognitive expenditure, particularly when both the verbal

and nonverbal channels of communication are being utilized. A key assumption shared by CAT, SFM and VOCSTAT is that all humans strive for stability and order in their interaction patterns. This communicative stability is what establishes the quality of communication that is influenced by all of the variables that Patterson describes.

VOCSTAT borrows from CAT the role of accommodation in mediating social relations and asymmetries, and also acknowledges the complex interplay between social variables and different modes of expression involved in SFM. The result is a model that predicts the importance of fundamental frequency in conveying social status relations.

Gregory and colleagues have illuminated the role that convergence plays in effective and efficient communication by using conversational analysis and long-term characterization of voice quality (as opposed to analysis of fine phonetic detail such as VOT, formants and pitch). Gregory and Hoyt (1982) analyzed taped conversations between Gregory and men stationed at an Air Force base for pause duration, pause frequency, turn-taking duration, turn-taking frequency and intensity. The results supported previous findings that conversational partners converge towards each other (Giles et al., 1973, 1977, 1979), but also showed that conversational pairs that exhibited the least amount of convergence also faced the most misunderstanding and miscommunication.

Gregory, Webster and Huang (1993) extended these findings to investigate convergence in fundamental frequency across conversational pairs. Data consisted of 12 more interviews between Gregory and men stationed at an Air Force base, as well as 11 interviews from an Egyptian Arabic database. They predicted that phonetic convergence would be detected in fundamental frequency and therefore used acoustic data that fell into the 62 to 192 Hz band, which corresponds to the F0 of most men. Listeners were asked to rate actual interviews (where convergence did occur) and virtual (pseudo) interviews (where convergence did not occur) for various qualities of the individuals involved in the interview and for the quality of the interview itself. Criteria consisted of 34 indices divided into evaluation, potency, and affect. Overall, their results supported their hypothesis; listeners determined that conversations were of a higher quality when convergence had occurred.

Another study conducted by Gregory et al. (1997) further demonstrated the importance of F0 in transmitting the social information that mediates convergence. Thirty pairs of talkers were placed in a room and asked to play a game that required them to verbally interact. The speech of one member of the pair was either high-pass filtered (where everything under 550 Hz was removed), low-pass filtered (where everything above 1000 Hz was removed) or unaltered. While unaltered speech was recorded from both members and examined for convergence, the other member heard only the high- or low-pass filtered speech in those conditions. Convergence was found in the high-pass and unaltered conditions, but not in the low-pass condition. Listeners were also asked to rate the quality of the conversation according to the 34 indices used in Gregory et al. (1993). Overall, they rated conversations that had been filtered lower than unaltered conversations, leading Gregory et al. to conclude that energy in the low frequencies of speech facilitates listeners' abilities to make judgments about social distance and accommodate accordingly.

Several reasons to account for the crucial role of F0 have been proposed. Firstly, the frequencies that correspond to F0 (usually about 60-300 Hz, depending on sex and individual anatomical variation) form part of the *tonal qualities* of speech that transmit information about the emotional state and "personality" of a talker (Kramer, 1964). Additionally, this range of frequency aids in the identification of talker through his or her non-content speech characteristics (Gregory et al., 1991). Finally, the absence of F0 may simply make a voice sound strange or unnatural, which may discourage talkers from identifying with it and subsequently accommodating to it.

However, their conclusion about the significance of F0 seems to ignore the fact that listeners can "fill in the blanks" in regards to energy in the lower frequencies by calculating F0 from harmonics in speech. If this is the case, the absence of F0 should not have such an effect on phonetic convergence. Babel (2009) suggests that the process of calculating the missing fundamental frequency creates "asynchronies in perceiving, processing and planning speech routines" and subsequently reduces talkers' abilities to converge. And while many other factors, such as gender and social role (Bilous and Kraus, 1988, Pardo, 2008) are involved in phonetic convergence, this study will seek to

further investigate the VOCSTAT model by having speakers participate in a shadowing task using both unfiltered and high-pass filtered speech.

III. *The "Shadowing" Paradigm*

Recent studies have utilized the immediate-repetition or "shadowing" paradigm employed by Goldinger (1998) to assess vocal imitation by having subjects hear and quickly repeat spoken words. In this experimental design, a baseline set of productions is collected first by having subjects read words off a computer screen, then exposing them to several blocks of shadowing in which they are asked to repeat the same words after a talker. Finally, a post-test set of productions is obtained by having subjects read the words again. Productions of shadowed words are then compared to baseline productions to determine whether any imitation occurred.

In a series of experiments utilizing this paradigm, Goldinger (1998) tested a model of episodic memory and speech perception designed by Hintzman (1986, 1988) called MINERVA 2. This model offers an exemplar-based account of convergence that differs markedly from the explanations offered by Giles and Gregory and their colleagues, which describe convergence as a social process about which individuals have at least some level of awareness and that reflects non-linguistic contexts such as status and dominance. MINERVA 2 offers an account of speech perception that assumes that every experience leaves an independent memory trace of perceptual and contextual information that is activated upon the presentation of a similar experience (or word). Every word that is presented causes an "analog probe" to all traces of the word that reside in long-term memory. Traces are activated in proportion to their similarity to the perceived token, and are collectively sent as an "echo" to working memory. Echoes differ in their intensity and their content. Echo intensity reflects "the total activity in memory" caused by the probe of a token (Goldinger 1998). A probe that is more similar to more existing traces will increase echo intensity. Echo content describes the "net response" of memory, or the content of the echo. A frequent token spoken in a familiar voice will strongly activate many traces, and if a match is perfect enough, a "generic echo," which will tend towards the mean of the set of activated tokens, will be sent to working memory. However, an infrequent word will generate a weak response by stored traces.

In immediate-repetition contexts, in which speakers repeat a word immediately after hearing a talker's production, the echoes sent to working memory strongly influence the production of the immediately-shadowed token. Therefore, the model predicts that the amount and strength of activation of existing traces generated by high-frequency word will augment the influence of memory on a shadowed production. Consequently, the influence of idiosyncratic qualities of a perceived token will be obscured by the strength of the echo in higher frequency words, and less convergence, or imitation, should occur. Goldinger confirmed this prediction in a series of shadowing tasks involving both English words and nonwords that had been assigned to high and low frequency conditions (1998).

IV. *Exemplar-based approaches to phonetic convergence*

In contrast to the non-exemplar based models discussed above, Goldinger's account does not consider convergence, or any variety of phonetic accommodation, to be a result of social processes. In exemplar-based accounts, convergence is an automatic process resulting from structure of the language system, or specifically from the interaction between speech perception and production. They posit that experiences, or traces, are actively stored in memory *during* speech perception and can affect later speech perception. For example, Palmeri, Goldinger, and Pisoni (1993) demonstrated that word-plus-voice information is stored upon perception. Listeners were continuously presented with old and new words in either the same voice or different voices, and asked to classify them as such. Correct classifications were higher for words presented in the same voice, even with delayed repetition or lag. This result suggests that perceptual details such as talker-specific information is encoded in the trace as well.

Additionally, conclusions derived from studies on non-linguistic imitation provide insight into the nature of spontaneous imitation. In primates, an automatic mental reflex activates mirror neurons whenever a behavior is observed (Gallesse and Goldman, 1998), and these mirror neurons in humans may play a role in both phonetic convergence and language learning (Babel 2009). However, Bargh and colleagues (Chartrand and Bargh 1996, Dijksterhuis and Bargh 2001, Bargh and Williams 2006) suggest that social factors can intervene in this reflex in higher-level primates and humans, inhibiting or enhancing imitation.

Other research supports the idea that convergence is not a purely automatic process. Recent studies conducted within Goldinger's shadowing paradigm demonstrate that even in immediate-repetition contexts, convergence can still be sensitive to social processes, namely indexical features such as gender and race. Namy et al. (2002) tested differences in vocal accommodation in men and women. Four talkers (2 men and 2 women) read a list of 20 words, which was subsequently shadowed by 8 males and 8 females. To further reveal any differences in perceptual sensitivities, 32 males and 32 females were asked to judge whether the baseline or shadowed productions were more similar to the talkers' own productions using a forced-choice AXB task (please see "Procedure" for a detailed description of this method). Results revealed that women converged to the talkers more than males, and that they converged more to male talkers. Males converged equally to male and female talkers. Additionally, the AXB task revealed that females detected convergence more often than males. This result supports another finding by Nygaard and Queen (2000) that women can more easily identify speakers using indexical features. Both sets of researchers concluded that gender differences in accommodation are at least partially driven by gender differences in perceptual sensitivity to indexical features. However, Namy does not deny the influence of social context on vocal accommodation tendencies, and suggests that the gender differences she observed may have arisen out of sociohistorically motivated differences in socialization. Women, she argues, are encouraged to attend more to indexical features such as intonation, tone of voice, and speaking rate, and may be more likely to incorporate them into their accommodation behavior.

Pardo (2006) also examined gender differences in phonetic convergence by having same-gender dyads participate in a map task. One member of the pair (the "giver") had a completed map that they used to direct their partner (the "receiver") towards completing their empty map. The task was designed to encourage participants to repeat the names of places on the map after each other. To assess convergence, Pardo also utilized an AXB task, which indicated that dyads converged 62% of the time. And, contrary to Namy's results, males were found to converge more often than females. Males converged more to givers, while women converged more to receivers. These

significant patterns reinforce the idea that particular social contexts can interfere in the perception-production link and facilitate convergence.

In her dissertation, Babel (2009) provides evidence that implicit social factors influence spontaneous phonetic imitation even within an exemplar-based model. She conducted a shadowing task using an Asocial Condition (where listeners only heard the talker's voice) and a Social Condition (where listeners additionally saw a photograph of the person they assumed to be the talker). See Appendix II for a detailed description of this study.

Babel's results suggest that phonetic convergence is "both phonologically and socially selective." Social effects entered into the process of phonetic accommodation through the influences of attractiveness ratings and IAT scores. For example, a social theory of accommodation (such as CAT) accurately predicts the behavior of female participants as a function of how attractive they found the talker: they imitated more in order to decrease the social distance between themselves and somebody they viewed positively. However, such theories don't explain why males would imitate more when they found the talker to be less attractive (although Babel suggests that asking mostly heterosexual males to rate another male based on physical appearance may have encouraged them to give false or inaccurate ratings out of discomfort).

Other ways in which social variables influenced the behavior of participants involved the difference in overall levels of convergence between the Asocial and Social conditions. Again, each gender exhibited opposing trends. Male participants converged significantly more in the Social condition, where a photograph accompanied the talker's voice, while the females converged more in the Asocial condition, although the trend was non-significant.

Babel, however, ultimately subscribes to an exemplar-based theory of convergence, but one in which implicit social values can enter into the perception-production link. Convergence proved to be automatic with respect to certain vowels, but implicit social factors, such as racial bias, had an inhibitory effect in most cases. An exemplar-based model is needed to account for such results, which indicated that memory encoded talker-specific variability allowed for implicit social measures to interfere.

It is clear that the study of phonetic accommodation needs to progress towards a more integrated theory that reasonably merges a social theory with an automatic one such that undeniable social effects are accounted for under an exemplar-based model. The current study represents an attempt to investigate the importance of social processes in phonetic convergence. It builds heavily on research on the significance of F0 conducted by Gregory and colleagues (see discussion above). The shadowing paradigm will be utilized in order to test the role of F0 in mediating phonetic convergence in an immediate-repetition task. Presumably, if F0 facilitates listeners' abilities to decode and process implicit social relations, its absence should inhibit phonetic convergence.

PROCEDURE

I. Participants

The model speaker: one adult male subject, who was a native speaker of American English provided all the stimuli for each shadowing task.

The Talkers: twenty-two native speakers of American English (7 males, 15 females) with no speech, language or hearing disorders served as participants for the shadowing task. The data for 3 subjects were unusable, due to either incompleteness or insufficient volume levels due to problems with audio equipment. In the end, 9 subjects (3 males, 6 females) completed the unfiltered condition of the Shadowing task, and 10 subjects (4 males, 6 females) completed the filtered.

The Listeners: twenty native speakers of American English formed the subject pool for the AXB task. Half (3 males, 7 females) were asked to judge the filtered condition, and half the unfiltered (4 males, 6 females).

All participants were recruited primarily from UC Berkeley undergraduate population and the Berkeley community through flyers, announcements and listings on websites such as the Research Subject Volunteer Program (RSVP). All participants were compensated $10/hour for their time, and were debriefed upon completion of their assigned task.

II. The Talker Task: Shadowing

The production list consisted of 39 English words (see Table 1). Words were not controlled for syllable count or for phonetic characteristics such as vowel quality since the AXB method of assessing convergence allows listeners the freedom of analyzing the similarity of words using any criteria available to them, such as those typically reserved for acoustic analysis, including F0, VOT, intonation and duration. However, it was noted that certain words included features such as flaps or unreleased word-final consonants that may be particularly salient to a listener assessing convergence in an AXB task.

hair	dog
yellow	people
everybody	discover
loud	cow
almost	hit
truck	house
together	terrible
please	airplane
pretty	picture
run	letter
story	remember
understand	question
park	book
bought	father
children	tomorrow
born	because
home	woman
apple	different
yesterday	better
fruit	

Table 1: Stimuli used in shadowing task.

[1]Originally, the list was designed for a shadowing task involving 9-10 year old children in order to investigate the influence of social role on phonetic accommodation in younger populations. While I eventually had to revise my purpose and protocol to eliminate the need for child subjects for the sake of time, the word list remained as one that was more appropriate for younger participants than for adult ones. Therefore, while

prior research
(Goldinger, 1998) has established that lower frequency words tend to exhibit the most
phonetic convergence, basic and relatively frequent words (as calculated by CELEX,
Baayen et al., 1993) were selected as target words in order to ensure that children would
recognize them. However, a few words with a frequency of 0 were included to assess the
impact, if any, of frequency on this particular experiment.

An adult male speaker recorded all of the audio-stimuli for the shadowing task in a sound-attenuated room. Using E-Prime Experimental Software (Schneider et al., 2002), the words in Table 1 were presented randomly on a computer screen four times each, and the speaker recorded his productions into a head-mounted AKG C520 microphone. The best tokens served as stimuli and were input into a shadowing paradigm task using E-Prime to be presented to subjects in the next part of the procedure.

The shadowing task was divided into two conditions, one using unaltered and unfiltered words and one using the same words that had been filtered twice at 300 Hz (see Appendix I: Figure 1 for sample spectrograms of a filtered and unfiltered token). Participants were native speakers of English with no speech, language, or hearing disorders. Upon arriving at the Phonology Laboratory at UC Berkeley, they were asked to review the word list for any questions and instructed to sit down at a computer workstation in a sound-attenuated room. Subjects wore a head-mounted AKG C520 microphone and AKG headphones. Their productions were recorded into the hard-drive of a PC. Instructions for the experiment were presented on a computer screen and subjects were left alone for the duration of the task.

The pre-task block of the experiment presented 39 words in random order in 36-point font on the computer screen. Subjects were instructed to read them "as naturally as possible," and were given two trial words for practice. In the three shadowing blocks that immediately followed, subjects were asked to repeat the word as naturally as possible upon hearing it over the headphones. Each block presented each word twice in random order, for a total of six shadowed repetitions of each word. 500 ms prior to the presentation of each word, the computer screen turned from white to red to alert the

subject that they were about to hear and repeat a new word. The final post-task block again involved simply reading the words off of the screen. The entire procedure lasted about 30 minutes.

III. The Listener Task: AXB

Individual tokens were extracted from each subjects' recording using a Praat script and input into an AXB task using E-prime, where the listeners were presented with 1). the baseline production of a Talker (Token A), 2). the original production of the same word by the model speaker (Token X) and 3). the shadowed production (Token B) taken from the same Talker's final block of shadowing. The order of presentation was also counterbalanced with the original Speaker's production always in the middle. Listeners were instructed to sit down in front of a computer in a quiet room and wear AKG headphones, and asked to determine whether the first two words (AX) or the last two words (XB) sounded more similar to one another by pressing one of two keys corresponding to either AX or XB.

The AXB task was divided into the filtered and unfiltered conditions, where half the Listeners judged words produced by Talkers in either condition. Both conditions of the task were divided into blocks corresponding to Talkers, so that the Listeners heard all 39 productions by one Talker before moving onto the next (twice, since the order was counterbalanced), with an opportunity for a short break in the middle. Order of presentation of the Talkers was randomized, as was the order of word presentation within each Talker. The entire procedure lasted about 90 minutes.

The task was restricted only to the comparison of the baseline production to the final shadowed production in order to keep its length manageable, but this decision unfortunately eliminated the opportunity to compare imitation between shadowing blocks. Imitation was judged to have occurred if the Listener judged the shadowed production as more similar to X than the baseline production, while divergence was judged to have occurred if the shadowed production was *less* similar to X. If my prediction about the role of F0 in accommodation holds, imitation will be detected more frequently in the unfiltered condition, in which the full range of speech is still intact and capable of communicating the social information through the F0 channel.

RESULTS AND DISCUSSION

I. The effect of condition

The AXB data were collapsed across conditions and filtered to include only responses with a reaction time (RT) between 300 and 4000 milliseconds, where the mean RT was 1638 ms. Outliers were defined as RTs that fell below or above that range. 5.0% of responses were classified as outliers based on RT, and were removed. An additional 75 responses were removed from Listener 103's final block of AXB judgments because she failed to complete the task, and the experimenter was obligated to finish for her. In total, 5.5% of responses were removed from the analysis.

Following previous analysis of similar data (i.e., Pardo, 2006), the data were subject to an ANOVA in order to test for the main effect of condition on how much imitation was detected. Imitation was defined as percent "correct" AXB judgments, where a "correct" judgment was scored whenever a Listener perceived that the shadowed token, as opposed to the baseline token, was more similar to the Speaker's production (X). Figure 1 demonstrates that a main effect of condition in the expected direction was found, where more imitation was detected in the unfiltered condition where F0 was left intact [$F(1, 13595) = 9$, $p < .01$] No effect of reaction time was found in the AXB task, suggesting that a significant delay (3000-4000 ms) between hearing the stimuli and judging them does not affect listeners' abilities to parse the fine phonetic detail contained in each token presented.

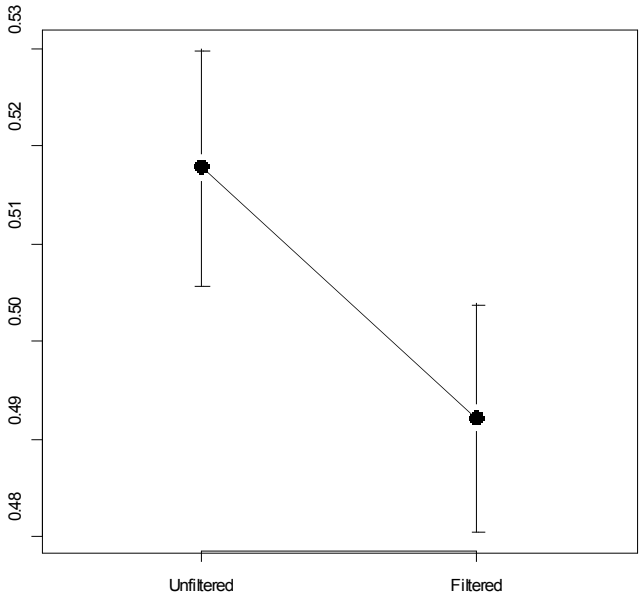

Figure 1: % Correct AXB judgments plotted for each condition. Talkers were judged to have converged 51.8% of the time when the Speaker's F0 was left intact, and 49.2% of the time when it was filtered out. While both percentages hover around chance, the difference in % correct AXB judgments between each condition suggests that Talkers were significantly more likely to imitate the Speaker if they were presented with the whole range of his speech.

This finding supports my prediction that the presence of F0 plays an important role in mediating phonetic convergence. Gregory and colleagues (Gregory, Webster and Huang, 1996; Gregory, Dagan and Webster, 1997; Gregory, Green, Carrothers and Dagan, 2001) conducted a variety of studies (described in detail above) that demonstrated that the presence of F0 facilitated interlocutor's abilities to assess the appropriate way to orient themselves within a conversational space (for instance, by converging to a speaker), but this study represents the first attempt to quantify its effect using an expansive collection of data from an immediate-repetition paradigm, as opposed to data gathered from natural conversation. Additionally, while Gregory et al., asked listeners to rate the "quality" of a conversation based on various indices, I asked listeners to directly judge, the extent of imitation by presenting them with a paradigm which asked them to compare tokens in isolation. And while there are a few methodological flaws in my procedure which will be discussed in depth below, it is encouraging and significant that Talkers did indeed converge significantly more to an "intact" voice than to one that has been obviously modified. However, it is rather unsurprising that talkers are less sensitive

and responsive to a voice that sounds less natural after having undergone high-pass filtering up to 300 Hz. This begins to address the question of whether talkers are simply not identifying socially with a "weird-sounding" voice, or whether they aren't capable of accommodating in relation to fundamental frequency, since F0 is absent from the speech they are hearing. Further research is necessary to determine whether listeners can calculate and "fill-in" F0 from the other harmonics and resonances in speech, and whether this process, as suggested by Babel (2009), impedes upon the interaction between perception and production which exemplar-based approaches to phonetic convergence use to account for imitation. Furthermore, it is still unclear if speakers who perceive a filtered voice as strange are in some way choosing *not* to implement their ability to determine and subsequently imitate F0. Perhaps one way to address this question would be to conduct a similar experiment, but using progressively more highly filtered speech, and asking participants to rate the "weirdness" of the voice. Presumably, an inverse relationship would exist between voices that rated high on the quality scale and the extent to which subjects imitated those voices.

In addition, Listeners in an AXB task have the freedom to judge similarity based on a multitude of acoustic factors besides F0, including duration, amplitude and intonation, so they may not have been paying attention to F0 at all, or doing so to variable extents, making it impossible to definitively describe changes in Talkers' F0 without acoustic analysis. In fact, Goldinger (1998) conducted a set of five AXB tasks in order to test for how important a number of acoustic factors were in cueing imitation. He generated four AXB tasks where word duration, amplitude, intonation contour and F0 were equated, respectively. Only tasks where duration and intonation contour were equated proved reliably different from the control task, where no aspect of the words was modified. This seemed to suggest, although somewhat inconclusively, that listeners pay most attention to temporal and melodic cues, not fundamental frequency. As mentioned earlier, acoustic analysis will be required to determine exactly what role or importance F0 played in my set of Talkers' judgments of imitation.

Due to this persistent uncertainty, this result, while important, does not contribute significantly to resolving the tension between automatic and social theories of accommodation. It is unclear whether, as Gregory et al. suggest, Talkers who do not hear

a Speaker's F0 are unable to detect his or her social or emotional status, and therefore cannot accommodate to those linguistic features, or whether the absence of F0 inhibits their ability to imitate the Speaker's lower frequency of speech (Gregory et al., 2001). Again, since AXB Listeners could have paid attention to a myriad of linguistic features when making their judgments, this question cannot be adequately answered without performing acoustic analysis.

Additionally, the explanation responsible for this observed effect of condition may actually rest, at least partially, in the perceptual systems of the AXB listeners who provided the judgments that determined to what extent imitation had occurred. When imitation was analyzed according to Talker, it was found that individual Talkers in both conditions differed widely in the degrees to which they imitated. Without acoustic analysis, it may seem difficult to determine whether their shadowed productions had indeed converged or diverged from the original token in terms of F0, or whether it was actually some other characteristic of their speech that impeded the ability of AXB Listeners to perceive imitation.

While fully assessing the validity of the first argument requires further data analysis, Figure 2 does provide evidence that objective phonetic accommodation did occur. One subject in the unfiltered condition (Talker 3) diverged from the model speaker, only imitating at a rate of 25.4% Correct AXB judgments. I have known the subject personally for many years, and without revealing too much background information, I can characterize her most succinctly as a "rebel," or someone who disdains conformism as well as following other people's examples. I believe that this prominent social characteristic accounts for the fact that she actually diverged from the Speaker as the shadowing task progressed. In this case, production most likely accounts for the Listener's judgments, which would probably be confirmed, at least to some extent, by

acoustic analysis.

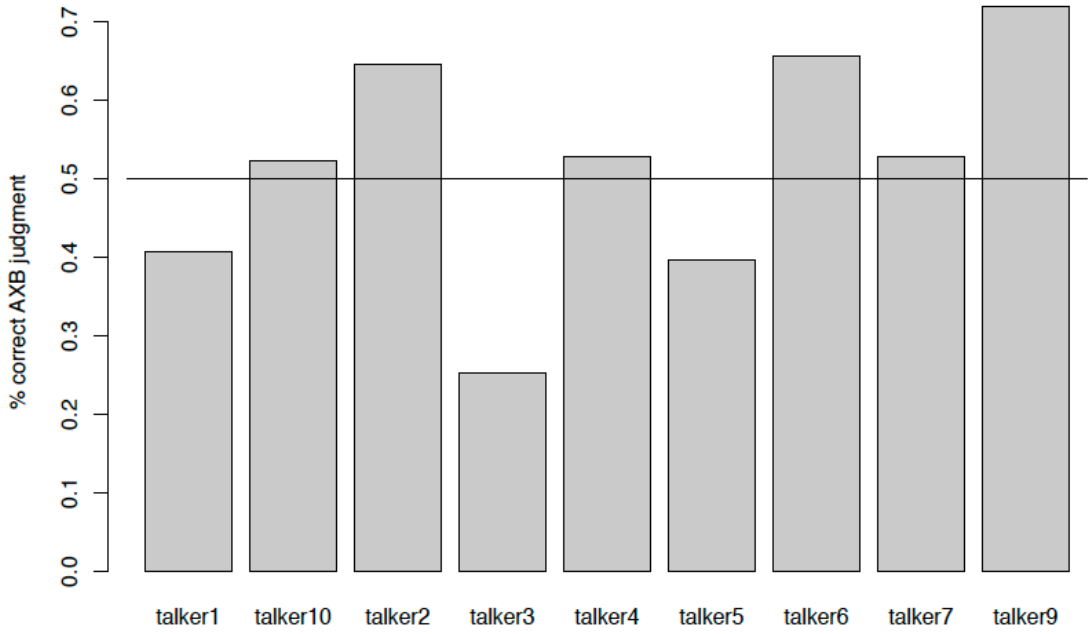

Figure 2: % Correct AXB judgments plotted for each Talker in the unfiltered shadowing condition. Variability exists between most Talkers, but most notably, Talker 3 diverges significantly from the Speaker's original productions, imitating only 25.4% of the time.

The results for the filtered condition (Figure 3), on the other hand, indicate the crucial role of the Listener in determining how much imitation was detected. Talkers 20 and 22 were judged consistently to have diverged at around 30% correct AXB judgments. Nothing about their language or personal backgrounds indicated that they were intentionally modifying their pronunciations of the target words to sound less similar to the original Speaker's (as in the case of Talker 3), but their language backgrounds did reveal that they were native speakers of Cantonese (Talker 20) and Korean (Talker 22), in addition to having learned English at ages 0 and 3, respectively.

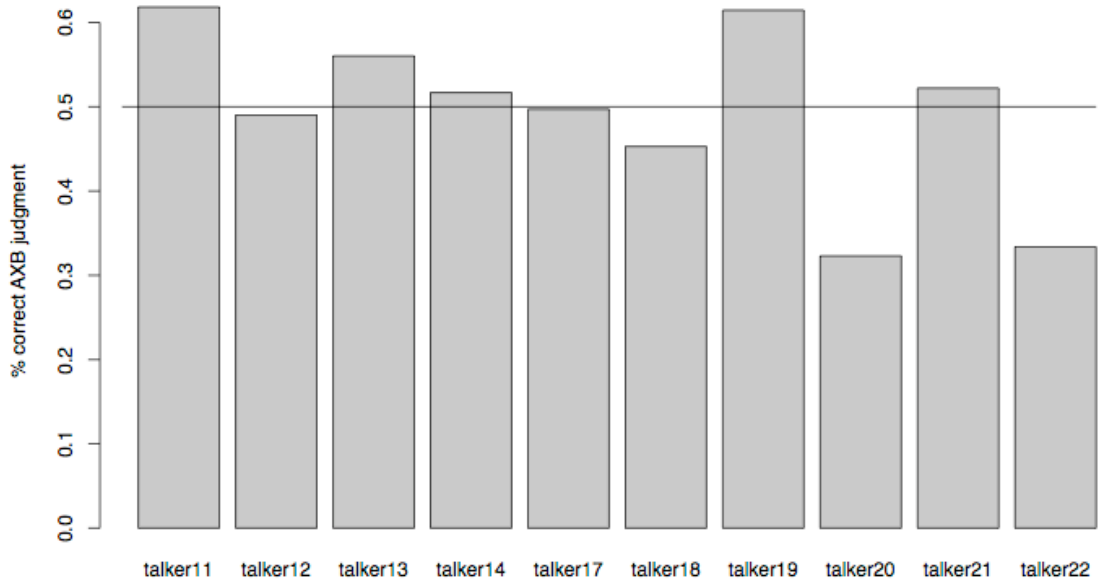

Figure 3: % Correct AXB judgments plotted for each Talker in the filtered condition. Talkers 20 and 22 were judged to have diverged, only imitating about 30% of the time each.

Additionally, in the debriefing portion of the AXB task, most Listeners mentioned that two of the voices that they heard sounded "Asian" (presumably these two voices were Talkers 20 and 22). Therefore, it is reasonable to assume that Talkers 20 and 22 spoke particular ethnolects that distinguished them quite significantly from the remainder of the Talkers who spoke more standard varieties of California English or Standard American English. Since none of the other Talkers possessed a noticeable ethnolect, the key to explaining this divergence most likely rests in the gap between the Listener's stored collection of standard variety exemplars, and the less-standard and less frequently heard tokens produced by Talkers 20 and 22. Perhaps Listeners were unable to detect any imitation in the productions of Talkers who used many distinct phonetic realizations (such as l-vocalization) because they were unused to judging small variations in dialects that were not representative of their own speech communities. To test this hypothesis, Listeners who speak the same ethnolects or dialects as these particular Talkers should be recruited to judge their productions against the original speaker's. Presumably, they will be more sensitive to variation and imitation and their judgments will match the direction and extent of accommodation that acoustic analysis would reveal. On the other hand, it is entirely possible that these Talkers did indeed diverge by emphasizing their phonetic or

dialectical differences over the course of the shadowing task, but again, this cannot be verified without in-depth acoustic analysis. However, if the Listeners were in fact in some sense "responsible" for this finding, there may be significant consequences for the study of dialect contact and formation. Paul Kerswill characterizes new variety formation as a product of both interactional, individual acts of accommodation that are responsive to particular contexts as well as of long-term accommodation that is sensitive to "wider, but community-specific social contexts" (2009: 3). Some important variables include the degree of contact between age and social groups, the types of relations between social groups and personal and social identities, as well as some demographic features such as the proportion of children to adults. However, it may be useful, in light of the current study, to include perceptual sensitivity to phonetic variants in the array of factors that lead to the creation of a new variety or to the modification of an existing one.

Another potential consequence of this study may apply to John Ohala's listener-based account of sound change that relies on the incorporation of a perceived new phonetic variant into a speaker's own production (Ohala, 1993). Perhaps there is a threshold of dissimilarity beyond which speakers are unable to adequately perceive changes in other varieties, and consequently do not respond to or incorporate those changes into their store of exemplars.

II. Talker and Listener Gender Interactions

Examining interactions between participant gender and behavior also yielded several interesting and significant results that suggest promising avenues for future research. Firstly, determining the behaviors of each gender within each task offered analyses with contradictory interpretations. Figure 4 shows that female Talkers were judged to have imitated significantly more frequently than male Talkers across conditions.

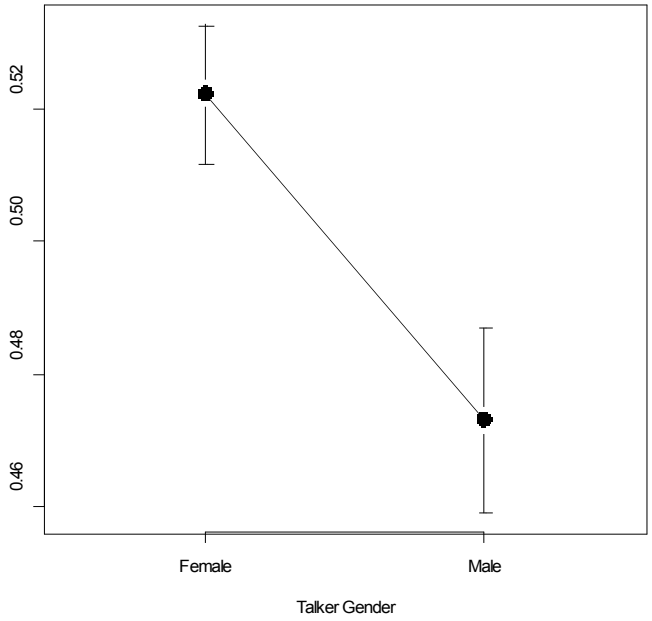

Figure 4: % Correct AXB judgments plotted for each Talker Gender across conditions, demonstrating that females imitated more than males (52.2% to 47.3%).

This finding is consistent with Namy and colleagues' result that female talkers in her shadowing task accommodated more than male shadowers (Namy et al., 2002). Drawing on previous research that demonstrated that females are better at identifying talkers than men (Nygaard and Queen, 2000), Namy et al., proposed that the reason female shadowers imitated more was due to gender differences in perceptual sensitivity or attention to indexical features that arise from different patterns of socialization. However, Pardo (2006) found the opposite result in her immediate-repetition study, and suggested that inherent differences in sensitivity play less of a role in imitation than the individual habitual attentional sets of both men and women.

The same analysis performed on Listener genders in the AXB task actually revealed the opposite pattern found in Talker Genders. Figure 5 plots % Correct AXB judgments against Listener gender across conditions and reveals that males actually detected more imitation overall.

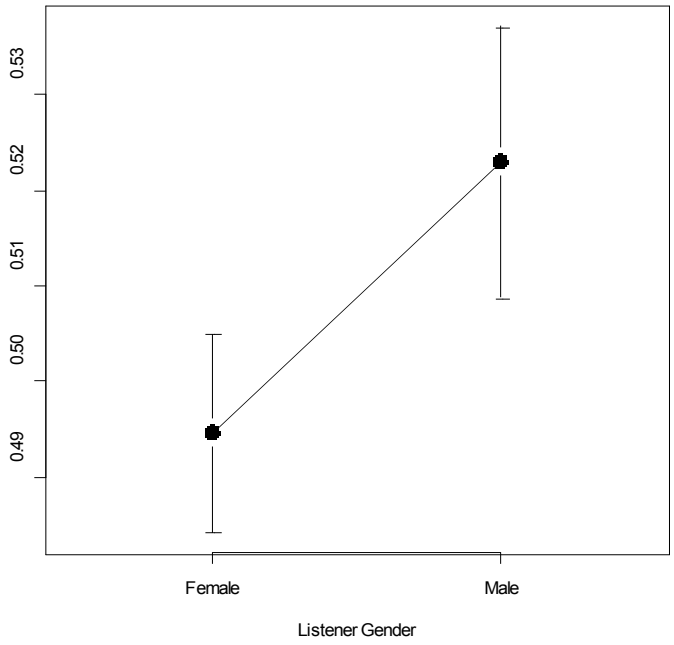

Figure 5: % Correct AXB judgments plotted for each Listener gender. Males were more likely to detect convergence than females (52.3% to 49.5%).

This finding lends support to Pardo's interpretation that individual attentional sets, regardless of gender, moderate apparent patterns of perceptual sensitivity, and suggests that differences in attention to indexical features correlated with differences in socialization do not play as significant a role as Namy et al. propose. That is, an individual's socialization experience shapes perceptual patterns more than differences in socialization across genders.

An investigation into the interaction between Listener Gender and % Correct AXB judgments across conditions also revealed opposing trends for male and female listeners. Females were more likely to detect imitation in the unfiltered condition than in the filtered, while men exhibited the opposite tendency (see Figure 6). Overall, however, males perceived more imitation than females across both conditions, as discussed above.

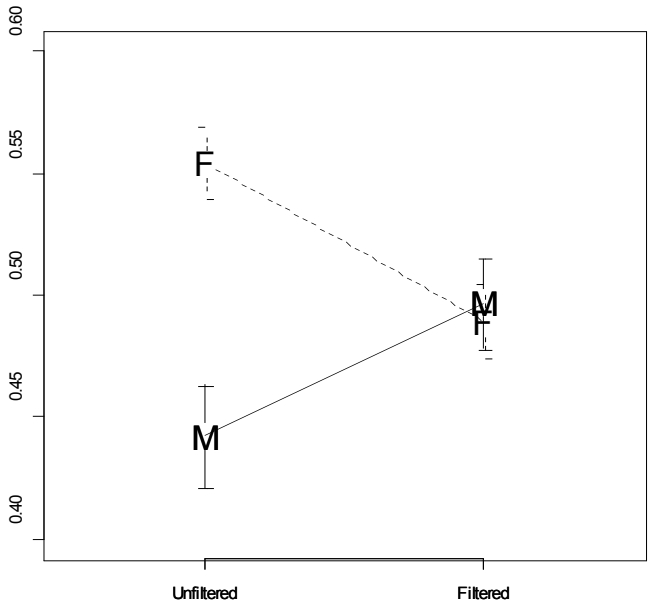

Figure 6: Females detected significantly more imitation than males in the unfiltered condition, although both genders converged to detect imitation at about the same rate (around chance) when F0 was filtered out.

It is unclear what can account for this observation. When presented with the full range of a speaker's voice, it may indeed be true that women are able to attend more carefully than men to indexical features such as F0, which conveys emotional information and plays a role in the dissemination of emotions. However, when women are confronted with filtered speech, they somehow lose that advantage in perceiving indexical features. Since F0 was most salient acoustic feature removed, this finding implies that its absence obfuscates the transmission of affective or emotional factors that women are socially encouraged to be sensitive to. However, the process by which this occurs is still unknown. Additionally, the reason why men were more likely to judge participants in the filtered condition to have imitated requires further research.

FUTURE DIRECTIONS

While this study provides some compelling cues into the way that speakers process and incorporate linguistic features of another speaker into their own productions,

even in very limited interaction, it raises numerous questions that require further research to resolve. While I have described several in the preceding sections, I will enumerate more here.

Firstly, the value of F0 was not calculated for each individual Talker, and the decision to filter out everything below 300 Hz was based on precedents set by prior researchers (Gregory et. al, 1997). A more precise methodology would call for taking into account individual variations in the range of F0, and filtered accordingly.

To eliminate the possibility that there were simply more Talkers whose voices were more accurately parseable for imitation in the unfiltered condition, the study should be repeated with more rigorous screening procedures (to ensure that all participants speak the same variety of English), and Talkers should not be divided into two conditions, but rather shadow both filtered and unfiltered speech from the same Speaker (perhaps in two separate sessions). Therefore, it would be possible to determine whether the effect of condition still held within each Talker, which would lend more evidence to the hypothesis that an individual's perceptual sensitivity and/or the ability to extract social information from a voice is inhibited when salient acoustic features are removed.

In addition, it would have been very useful to ask listeners to judge tokens taken from all blocks of shadowing in order to track imitation across blocks. This would illuminate the progress of accommodation within an interaction. However, this all-inclusive task would probably require an inordinately long time commitment from participants (five hours, in this particular study), and it would be more feasible to break up the task and recruit more listeners, a decision that relies on the assumption, which has been disconfirmed by this study, that all listeners detect imitation at about the same rate.

Finally, this data may have implications for the study of how speakers perceptually group other talkers by styles, dialects and ethnolects. Ongoing acoustic analysis may elucidate the link between short-term phonetic accommodation and long-term accommodation that results in the propagation of sound change and the formation of new dialects. The author and Dr. Babel are currently running this analysis.

References

Babel, Molly. 2009. Phonetic and social selectivity in speech accommodation. Doctoral dissertation, University of California, Berkeley.

Baayen, R.H., R. Piepenbrock, and H. van Rijn. 1993. The CELEX lexical database CDROM). Linguistic Data Consortium, University of Pennsylvania.

Bilous, Francis R., and Robert M. Krauss. 1988. Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. *Language & Communication* 8:183–194.

Bourhis, Richard Y., and Howard Giles. 1977. The language of intergroup distinctiveness. In *Language, ethnicity, and intergroup relations*, ed. Howard Giles, 119–136.

Burgoon, J. K., Le Poire, B. A., & Rosenthal, R. (1995). Effects of preinteraction expectancies and target communication on perceiver reciprocity and compensation in dyadic interaction. *Journal of Experimental Social Psychology*, 31, 287-321.

Chartrand, Tanya L., and John A. Bargh. 1996. The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology* 76:893–910.

Dijksterhuis, Ap, and John A. Bargh. 2001. The perception-behavior expressway: Automatic effects of social perception on social behavior. In *Advances in Experimental Social Psychology*, ed. Mark Zanna, 1–40.

Bargh, John A., and Erin L. Williams. 2006. The automaticity of social life. *Current Directions in Psychological Science* 15:1–4.

Giles, Howard. 1973. Accent mobility: a model and some data. *Anthropological Linguistics* 15:87–105.

Giles, Howard, and Nikolas Coupland. 1991. *Language: Contexts and consequences*. Milton Keynes: Open University Press.

Giles, Howard, Nikolas Coupland, and Justine Coupland. 1991. Accommodation theory: Communication, context, and consequence. In Contexts of Accommodation, ed. Howard Giles, Justine Coupland, and Nikolas Coupland, 1–68.

Giles, Howard, Donald M. Taylor, and Richard Bourhis. 1973. Towards a theory of interpersonal accommodation through language: some Canadian data. *Language in Society* 2:177–192.

Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105:251–279.

Greenwald, Anthony G., Debbie E. McGhee, and Jordan L. K. Schwartz. 1998. Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology* 74:1464–1480.

Gregory, Stanford W., Kelly Dagan, and Stephen Webster. 1997. Evaluating the relation between vocal accommodation in conversational partners' fundamental frequencies to perceptions of communication quality. *Journal of Nonverbal Behavior* 21:23–43.

Gregory, Stanford W., Brian E. Green, Robert M. Carrothers, and Kelly A. Dagan. 2001.

Verifying the primacy of voice fundamental frequency in social status accommodation. *Language and Communication* 21:37–60.

Gregory, Stanford W., and Brian R. Hoyt. 1982. Conversation partner mutual adaptation as demonstrated by fourier series analysis. *Journal of Psycholinguistic Research* 11:35–46.

Gregory, Stanford W., Stephen Webster, and Gang Huang. 1993. Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. Language and Communication 13:195–217.

Gregory, Stanford W., and Stephen Webster. 1996. A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology* 70:1231–1240.

Hintzman, D.L. 1986. "Schema abstraction" in a multiple-trace memory model. *Psychological Review* 93:411-428.

Hintzman, D.L. 1988. Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review* 95:528-551.

Kerswill, Paul. 2009. Contact and new varieties. Forthcoming, *Handbook of Language Contact.* Oxford: Blackwell.

Kramer, E. 1964. Elimination of verbal cues in judgments of emotion from voice. *Journal of Abnormal and Social Psychology*, 68:390-396.

Namy, Laura L., Lynne C. Nygaard, and Denise Sauterberg. 2002. Gender differences in vocal accommodation: the role of perception. *Journal of Language and Social Psychology* 30 21:422–432.

Nygaard, L.C., & Queen, J.S. 2000. Surface form typicality and asymmetries in recognition memory. *Journal of Experimental Psychology: Learning, Memory & Cognition* 26: 1228-1244.

Ohala, John J. 1993. Sound change as nature's speech perception experiment. *Speech Communication* 13: 155-161.

Ohala, John J. 1981. The listener as a source of sound change. *Papers from the Parasession on Language and Behavior.* Chicago Linguistics Society 178-203.

Palmeri, Thomas J., Stephen D. Goldinger, and David B. Pisoni. 1993. Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19:309–328.

Pardo, Jennifer S. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119:2382–2393.

Pardo, Jennifer S. 2008. Acoustic-phonetic variation and conversational interaction. *In Expressing Oneself/Expressing One's self: A Festschrift in Honor of Robert M. Krauss*, ed. Ezequiel Morsella, in press.

Tajfel, Henri and Turner, John.1979. An integrative theory of intergroup conflict. *The Social Psychology of Intergroup Relations*. 94-109.

Yang, Kuo-Shu and Michael H. Bond. 1980. Ethnic affirmation by Chinese bilinguals. *Journal of Cross-Cultural Psychology* 11:411-424.

APPENDIX I

Figure 1:

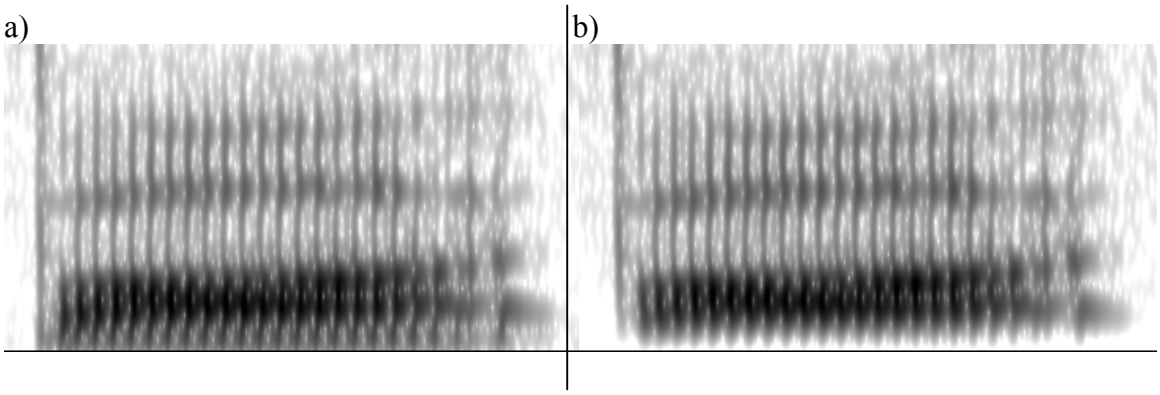

Figure 1: Spectrograms for stimulus "bought." a). is unfiltered speech, b). has undergone high-pass filtering at 300 Hz.

APPENDIX II

Two talkers, one Black and one White, were used in Babel's study. One hundred seventy-eight participants were assigned relatively equally to six conditions created by a fully-crossed experimental design:

1. Black Asocial Condition: Black talker voice, no picture.
2. White Asocial Condition: White talker voice, no picture.
3. Black Social Condition: Black talker voice, Black talker picture.
4. White Social Condition: White talker voice, White talker picture.
5. Black/White Social Condition: Black talker voice, White talker picture.
6. White/Black Social Condition: White talker voice, Black talker picture.

However, as only 22 Black participants completed the task successfully, the data only included 154 White participants. Babel also asked each participant in the Social conditions to rate the attractiveness of the model talker and asked them to complete an Implicit Association Task (Greenwald et al., 1998) in an effort to determine how implicit biases might interact with cognitive processes such as phonetic convergence by measuring how such biases prime behavior in the shadowing task.

The stimuli for the task consisted for 20 Black and 20 White names, along with 20 semantically "good" and 20 semantically "bad" words. The task itself was comprised of 5 blocks and was presented on a computer. The rightmost and leftmost buttons on a button box were used by participants to categorize. The first asked participants to characterize a name that was shown in the middle of the screen as either BLACK or WHITE, which were written in either top corner. Participants were instructed to press the corresponding button (right or left) to categorize the names. Secondly, participants had to categorize semantically "good" or "bad" words (e.g. diamond or awful) according to the attributes *good* and *bad*, which were presented in the top corners of the screen in place of targets BLACK or WHITE. In the third block, both target-concepts and attributes were presented in the top corners, and either names or words appeared in the middle of the screen. Participants were asked to classify either names or words into their corresponding categories. The fourth block was identical to Block 1 except that BLACK and WHITE switched corners. Finally, the fifth block corresponded to the third block, except that the switched order of BLACK and WHITE was paired with the original order of *good* and *bad*.

Individuals' performance on the task was scored by calculating a mean based on the correct responses for each block, which along with a 600 ms penalty, replaced each mean response error. One standard deviation was found for each block. The means for each block were then re-calculated and the difference was determined and divided by the one standard deviation to get the IAT score. A negative IAT score indicated a pro-Black bias whereas a positive IAT score indicates a pro-White bias.

The data revealed that male participants who scored with less of a pro-White than others were more likely to imitate the Black talker. IAT ratings had a much less significant correspondence on female participants, the effect of attractiveness interacted with convergence in such a way that females were more likely to imitate a talker if they found him more attractive. Males and females, however, imitated to the same overall degree, even though male participants had the physiological ability to more closely imitate the actual phonetic characteristics of the male talkers.

APPENDIX III

ACKNOWLEDGMENTS

I would like to express my gratitude to all of my professors, fellow students, and friends, whose support and encouragement made this thesis possible. In particular, I would like to thank Professor Keith Johnson for his invaluable advice and support throughout this process. In addition, I am thankful to Ron Sprouse for always being available to help with lab equipment and programs.

I would like to thank the Department of Linguistics, who not only provided allowed me the use of necessary equipment and resources, but who opened the opportunity to conduct research. It is an honor to have learned and worked within such a world-class program.

I appreciate the kindness of my classmates and friends, who expressed interest in my research and who volunteered their time in order to participate.

Finally, I offer my sincerest gratitude to Molly Babel, whose guidance and endless patience made this study possible.