Selective vowel imitation in spontaneous phonetic accommodation

Molly Babel
Department of Linguistics
University of British Columbia and University of California, Berkeley

I. INTRODUCTION

As people learn to speak, they acquire the language and dialect spoken around them. Sentence structure, lexical selection, and pronunciation are all determined by the patterns used in the ambient language. With respect to pronunciation, this pattern of sounding like those in our immediate environment continues throughout the lifespan through the process of *spontaneous phonetic imitation* (Goldinger, 1998). This is the phenomenon in which a talker acquires the acoustic characteristics of an interlocutor through the course of verbal interaction. Spontaneous phonetic imitation is important, since it may account for a wide range of phenomena such as historical sound change and dialect acquisition. Moreover, the cognitive mechanisms that motivate spontaneous imitation are pervasive from infancy (Kuhl and Metzloff, 1996) and existent across non-linguistic behaviors like foot shaking and face touching (Dijksterhuis and Bargh, 2001). In terms of speech, the cognitive mechanisms that prompt actions that are taken in by the perceptual field are of great importance for theories of speech perception and production, crucially for theories that propose a strong link between the two faculties (e.g., Goldstein and Fowler, 2003).

Phonetic imitation, also known as phonetic convergence or phonetic accommodation, has implications for variability in language representation as well. The fact that the perceptual categories of language are labile is well grounded (Norris et al., 2003; Kraljic and Samuel, 2005, 2006, 2007; Kraljic et al., 2008a,b; Maye et al., 2008).

In terms of speech production, between-talker variability can often be explained by physiological differences between talkers. However, if psychological differences between talkers are ignored, a large part of how language users acquire the local parlance involves studying the malleability of production, which in turn can be investigated through phonetic imitation.

Laboratory research has revealed that phonetic convergence can occur in both socially minimal situations where talkers are simply producing single words (Goldinger, 1997, 1998; Goldinger and Azuma, 2004; Namy et al., 2002) and in cooperative, socially rich, dyadic interactions (Pardo, 2006, in press). Scholars exploring issues of dialect contact have also engaged in research on phonetic accommodation, as dialect acquisition can roughly be considered a long-term exercise in speech accommodation. In most cases, over time a talker will begin to sound more like the talkers of the new dialect, losing elements of the original dialect. With a few exceptions (Shockley et al., 2004; Evans and Iverson, 2007; Nielsen, 2008; Delvaux and Soquet, 2008), studies of phonetic imitation or dialect acquisition have ignored linguistically meaningful phonetic content. Instead previous work has focused on broad acoustic measures and non-contrastive phonetic traits like pitch or duration (with American English talkers) or they have used perceptual similarity judgments from naïve listeners as the metric with which to determine the presence of phonetic imitation.

The experiment described below addresses this gap in the knowledge of what is imitated by specifically reporting on how participants imitate vowel spectra in a lexical shadowing task. Crucially, this experiment provides a nice test-case for exemplar-based theories of speech (Johnson, 1997; Goldinger, 1997; Pierrehumbert, 2001, 2003). A

simple description of such a theory is as follows. Upon hearing a word, episodic traces associated by the talker voice or word are activated in memory. The more familiar a voice or the higher frequency the word, the higher the number of activated traces because of increased experience. Goldinger (1997:46) clarifies, "even if an exact match to the [word] exists in memory, all similar activated traces create a 'generic echo,' regressing toward the mean of the activated set." It is the mean of the activated set that is selected for production. Goldinger's theory predicts that upon perceiving a particular word, the activated traces will contribute to its production. Thus, in terms of phonetic convergence, exposure to words produced by a model talker will shift a participant's productions toward those of the model talker, but only if there are pre-existing traces in the direction of the model talker token. So, for example, in the task described below words contain the vowels /i æ ɑ o u/. If a model talker produces these vowels with overall lower F1 than a participant, the participant's traces of these vowels with a low first formant would be activated. Because talkers have an increased number of variants within the first formant dimension for low vowels because of differences in jaw height (Summers, 1987), it would be predicted that the participant would accommodate to the lower F1 in only the low vowels /æ/ and /ɑ/. If the activation of traces has no bearing on which vowels are imitated, then all vowels should be accommodated to the same degree.

Before describing the experiment, previous work on phonetic accommodation across dialects and within dialects are reviewed. These summaries are followed by the results of the lexical shadowing task and a discussion of the findings.

A. Across dialect

When a talker moves into a new dialect area, they gradually acquire many of the phonetic characteristics of the new dialect. While young children acquire language with ease, it has been documented that after a certain age children are unable to fully acquire particularly complex phonological aspects of a second dialect (Payne, 1980; Chambers, 1992). Trudgill (1986:58) argues this is due to the fact that accommodation of sound structure in dialect contact is "*phonetic* rather than phonological" (emphasis in original); in other words, according to Trudgill, talkers target particular words and not entire phonological patterns when they accommodate. In the dialect contact literature, the accommodation processes and the mechanisms that motivate them are under intense discussion (cf. Trudgill, 2008).

Phoneticians and laboratory phonologists have also examined dialect change. In a study of dialect change of Canadians living in Alabama, Munro et al. (1999) found that Canadians living in Canada perceived the speech of Canadians living in Alabama to be somewhere between that of Canadians in Canada and native Alabamans. While there was considerable individual variation among the Canadians in Alabama that was somewhat related to the length of residence, their speech had generally accommodated to the speech of Alabamans.

In a longitudinal study, both production and perception in new dialect exposure were examined by Evans and Iverson (2007). Nineteen college students from a northern British English dialect were interviewed four times throughout the first two years of study at a southern English university where Standard Southern British English (SSBE) was spoken. Of particular interest in this research was the behavior of the vowels in *bud*,

cud, and *bath*. SSBE has /ʌ/ in *bud* and *cud* and /ɑ/ in *bath*. Northern varieties of English use the vowel /ʊ/ for *bud* and *cud* and /ɑ/ for *bath*. Acoustic analyses revealed that *bud* and *cud* became more centralized. In the northern dialects, *cud* and *could* are homophonous; both have the vowel /ʊ/. *Could* has this vowel in the southern dialects, but *cud* has /ʌ/. After their time at the university, the participants began to centralize the vowel for *could* as well. Participants were also rated on a ten-point scale from 'very northern' to 'very southern'. Overall participants were rated as sounding more southern after their time at the university. Despite these changes in production, there was no overall change in the position of participants' perception based on an examination of their best exemplar locations, although participants who were rated as having maintained a more northern accent did in fact choose more northern exemplars.

An important recent study involving imitation across dialect is Delvaux and Soquet (2008). Delvaux and Soquet examined shifts by French female speakers of the Mons dialect of Belgian French (n = 8) to the Liège dialect of a Belgian French female model talker. Duration and mel frequency cepstral coefficients (MFCC) for /o/ and /ɛ/ – two vowels that differ across the two regiolects of Belgian French – were examined. In these two dialects of Belgian French the formant values of the mid-back vowel vary ([o]~[ɔ]) and the duration of /ɛ/ varies ([ɛ]~[ɛ:]). Target words containing the key phonemes were elicited in a sentence production task. Mons speakers imitated the model talker based on the MFCC measures and duration of both /o/ and /ɛ/. The effect

diminished but still persisted into a post-task production test, indicating exposure to the regiolect had a long-standing effect on the representation of the vowels.

B. Within dialect

Only recently have studies of accommodation and imitation begun to explore the segmental effects of phonetic imitation (Shockley et al., 2004; Nielsen, 2008; Pardo, in press). The earliest studies examined broader acoustic measures and relied more on perceptual judgments from listeners. Natale (1975a,b) found that in dyads interviewees accommodated to intensity levels and temporal patterns of the interviewer. In a series of studies, Gregory and colleagues examined behaviors across conversational dyads. They found that conversational partners converge with respect to intensity, pause duration, pause frequency, turn-taking duration, and turn-taking frequency (Gregory and Hoyt, 1982) and long-term average spectra (Gregory et al., 1993; Gregory and Webster, 1996; Gregory et al., 1997; Gregory et al., 2001).

Goldinger (1998) examined phonetic imitation at the word level using a lexical shadowing task to prompt imitation and an AXB task to elicit judgments of perceptual similarity (i.e., imitation). Goldinger found lexical frequency, amount of exposure, and shadowing time (immediate or delayed) to interact with degree of imitation. In the immediate shadowing condition, all words were judged to have been imitated for all talkers with increasing word frequency inhibiting convergence. In the delayed shadowing condition only low and middle-low frequency words showed convergent behavior. To determine what aspects of the acoustic signal were converging, Goldinger (1997) reported the results of a pilot study involving acoustic measurements. Average

pitch, intonation, and word duration were measured. All participants deviated from their baseline average pitch toward the pitch of the talker they were exposed to in the shadowing task. In another task, Goldinger and Azuma (2004) revealed that effects of episodic memory traces are activated in orthographic representations as well. Goldinger's (1998) basic finding was replicated by Namy et al. (2002); however they expanded on the result by discovering that female participants accommodated to the model talkers more than male participants.

Some of the most influential recent work on phonetic convergence is reported in Pardo (2006). Pardo examines phonetic convergence in same-gender dyads involved in jointly completing a map task. Pardo also employed Goldinger's (1998) paradigm; the X tokens in the AXB similarity task were taken from the recordings of an individual's conversational partner in the map task. Dyads were perceived to have converged on 62% of the trials. Listeners judged that male dyads converged more than women (75% vs. 58%). Women were found to converge toward the speaker who was receiving instructions. Men patterned oppositely; they converged toward the speech of the male talkers giving instructions.

Pardo (in press) reports on the acoustic analyses conducted on the speech of the same-gender dyads alone, and also presents results from mixed-gender dyads. Like Pardo's same-gender dyads, mixed-gender pairs converged, albeit to a lesser extent; listeners perceived mixed-gender pairs to have converged on 53% of the trials. Pardo conducted linear regressions with F0 and duration data to determine the cues on which listeners based their judgments. These values accounted for 41% of the variance for the female talkers, but only 7% of the variance for the male data.

This research has established that talkers do spontaneously imitate and accommodate in verbal interaction, but less is known about *what* within the acoustic signal is the target of imitation. In order to understand the affect of imitation on language change and dialect acquisition, scholars should be concerned with the phonetic details involved in speech accommodation. Along these lines, some work has been done on the imitation of lengthened VOT. Shockley et al. (2004) demonstrated that talkers imitate artificially lengthened VOTs in American English aspirated stops. Nielsen (2008) further expanded on VOT imitation in American English, finding an interaction between imitation and generalization across the system. Nielsen revealed that not only do participants imitate the increased VOTs for words they are exposed to (in this case, /p/-initial words), but they generalize the lengthened pattern to a segment sharing the same [+ spread glottis] feature (/k/-initial words). A word-specificity effect was found as the increased VOT effects are strongest in the /p/-initial words heard in the exposure phase. Another group of listeners was presented with /p/-initial words with shortened VOTs. Listeners in this group did not imitate the shortened /p/ VOTs, nor did they apply any such pattern to the /k/-initial words introduced in the test phase. Nielsen reasons that the shortened VOTs encroach on the phonetic space of the unaspirated stops in American English.

The effect of language knowledge and phonetic/phonological space on imitation is also investigated in Nye and Fowler (2003). Participants shadowed sentences that varied in terms of their phonotactic approximation to English. Shadowing was more reliable when the sentences were more English-like, suggesting that language-specific phonotactic knowledge guides imitative faithfulness and accuracy.

Recently, Mitterer and Ernestus (2008) argue that imitation is phonological (and not phonetic) using data from a speeded shadowing task in Dutch. In the experiment participants were presented with either an alveolar or uvular variant of /r/. The group of participants was equally split between individuals who were habitual users of the alveolar trill and those who were habitual uvular trill users. In general, participants retained their habitual pronunciations and did not imitate the place of articulation. Other stimuli had either a voiced or unaspirated stop in onset position. The voiced stops differed with respect to the number of pre-voicing cycles. Participants imitated the presence or absence of pre-voicing, but the difference in amount of pre-voicing was not imitated. Mitterer and Ernestus conclude from these results that the relationship between speech perception and production is linked by abstract phonological representations.

Previous work has established that users of language accommodate and imitate each other phonetically, but the knowledge about the aspects of the phonetic signal that undergo imitation are lacking. In terms of aspirated American English stops, the work by Shockley et al. (2004) and Nielsen (2008) has established that VOT is imitated. Evans and Iverson (2007) demonstrated that words inhabiting one vowel category can shift after exposure to new dialect. Delvaux and Soquet (2008) also found that talkers modify vowel productions after exposure to a slightly different dialect. The purpose of the experiment reported here was determine whether talkers specifically imitate vowel spectra in a lexical shadowing task.

II. METHODOLOGY

A. Stimuli

Fifty low frequency monosyllablic words from CELEX (Baayen et al., 1993) with the vowels /i æ ɑ o u/ were selected as stimuli (see Table 1). Low frequency words were selected as Goldinger (1998) reports that low frequency words were imitated more than high frequency words. Two male participants served as model talkers for the experiment. Both talkers worked in the same office at the University of California, Berkeley and were in their early thirties. One talker was White and one was Black. To the author's ear, both talkers spoke a standard variety of California English. Neither talker had speech, language, or hearing disorders. They were both compensated $10 for their time.

/i/	/æ/	/ɑ/	/o/	/u/
breeze	bask	clock	close	bloom
cheek	bat	clot	coat	boot
deed	mask	cot	comb	doom
freak	nag	pod	foal	dune
key	smash	sock	hone	glue
peel	snap	sod	mote	hoop
sneeze	tap	spawn	soap	pool
teal	vat	stock	toad	tool
teethe	wag	tot	tone	toot
weave	wax	wad	woe	zoo

Table 1. Stimuli used in shadowing task. All words have raw frequency counts of ≤1 per million in spontaneous speech in the CELEX database.

The audio-stimuli for the experiment were recorded in a sound-insulated booth with a head-mounted AKG C520 microphone positioned three inches from the talker's mouth. Stimuli were randomly presented on the computer monitor four times each through E-prime Experimental Software (Schneider et al. 2002). The most natural

and clear sounding of the four tokens of each word as determined by the author was selected for use in the experiment. The talkers were also digitally photographed for the visual stimuli.

B. Participants

One hundred and seventeen participants (male = 53) completed the task. The data from four participants were excluded from the analysis. Two of these participants were removed because they did not complete the task accurately in the sense that they did not produce the words naturally. Another was removed because it was discovered after the task that she had profound hearing loss. The final participant was removed because she personally knew the model talker to which she had been assigned. The remaining participants have no reported speech, hearing, or language disorders. All participants were compensated $10 for their time.

C. Procedure

Participants in the shadowing task were tested individually and were randomly assigned to one of four conditions. Details about the experimental conditions are summarized in Table 2. The paradigm for the speech production task is a lexical shadowing paradigm (Goldinger, 1998; Namy et al., 2002). The procedure in all four conditions was identical. Participants were seated in a sound-attenuated room at a computer workstation where the experiment was presented using E-Prime Experimental software (Schneider et al., 2002). Participants wore a head-mounted AKG C520 microphone positioned about 2 inches to the side of the mouth and AKG K240 headphones. Word productions were digitally recorded to the hard drive of a PC at a 44K sampling rate. Recordings were down sampled to 22K before analysis.

	Voice	Picture
Black talker with no visual prompt	Black	none
Black talker with visual prompt	Black	Black
White talker with no visual prompt	White	none
White talker with visual prompt	White	White

Table 2. The four experimental conditions and their design. Voice refers to the racial identity of the talker whose voice was used in the condition. Picture refers to whether there was a digital image of the talker presented in the condition.

The task proceeded as follows: The first block was a pre-task block to establish participants' baseline productions of the words. In this block, words were randomly presented one time each in 36-point font in the middle of the screen. Participants were instructed to read the words as naturally and clearly as possible. In the test blocks, the randomized word list was presented binaurally at 65 dB (SPL) over the headphones. The test blocks were comprised of three shadowing blocks where words were repeated twice per block for a total of six repetitions of each word. Each trial began with the screen turning from white to red 500 ms before the presentation of an audio file. Participants were informed that upon hearing the word, they were to repeat it as clearly and naturally as possible. In the Visual Prompt Conditions a talker photo (sized 466 x 366 pixels) was presented on the screen for the duration of the shadowing portion of the task. Finally, the post-test block was identical to the baseline block where participants read the words from the screen. The experiment took less than half an hour to complete. Upon exiting the sound booth participants in the No Visual Prompt Conditions were asked to identify the race of the model talker. Participants did not reliably identify the face of the Black talker; both the Black and White talker were both identified as White by participants (χ (1) = 0.08, p = n.s.).

D. Data Analysis

A Praat (Boersma and Weenink, 2005) script automatically identified pauses by marking boundaries preceding and following over 600 ms of low intensity energy (< 59 dB). These boundaries were hand-corrected so as to mark the onset and offset of the vowel. A second script extracted the mean first and second formants from a series of Gaussian windows spanning the middle 50% of the vowel with a 2.5 ms step size. Outliers were identified as those tokens where the F1 or F2 was more than three standard deviations away from the mean. This was based on the group mean for each formant for each vowel and was done separately for the male and female data. Outliers were then removed from the dataset. As per Adank et al. (2004), the Lobanov normalization method was used (Lobanov 1971).

The metric of interest is how participants' productions change as a result of auditory exposure to the model talker. As a measure of how productions changed, a measure of how much the production of a particular word changes through the course of the experiment is needed. To this end, the Euclidean distance was calculated from each participant production to the model talker production of the same word from the talker to which they had been assigned. From the set of distance calculations there is a measure of the acoustic distance between the model talkers' productions and the participants' productions (400 productions per participant). To calculate how much a participant modified their production as a result of being exposed to the model talker a comparison between the original distance of a participant's baseline productions and those of the shadowing and post-task blocks was necessary. Therefore, the original baseline distance for each word was subtracted from the distance for each following instance of that word.

The value calculated from each instance of subtraction is the *difference in distance*. A negative difference in distance value demonstrates that the phonetic distance between the participant and the model talker shrank and that some degree of phonetic accommodation took place. A positive value indicates an increase in phonetic distance (i.e., divergence). A value of 0 demonstrates that there was no change as the result of auditory exposure to the model talker. This difference in distance value is used as the dependent measure in the statistical analysis.

III. ANALYSIS AND RESULTS

The full design of the experiment was a 2 (Voice: Black talker or White talker) x 2 (Picture: Visual Prompt or No Visual Prompt) x 5 (Vowel: i æ ɑ o u) x 4 (Block: Shadowing Blocks 4, 5, 6 and Post-task Block 7) x 2 (Gender: Male or Female) factorial design. In this design, Vowel and Block were within subject variables. Difference in distance values were summarized across cells and the means were used in a repeated measures analysis of variance. Vowel [$F(4, 396) = 56.2$, $p < 0.001$] and Block [$F(3, 297) = 60.7$, $p < 0.001$] returned as main effects. These factors also were significant as a two-way Vowel x Block interaction [$F(12, 1188) = 20.2$, $p < 0.001$] and in a three-way interaction of Vowel x Block x Talker Race [$F(12, 1188) = 3.8$, $p < 0.001$]. There were also two-way interactions with Vowel x Gender [$F(4, 396) = 2.4$, $p < 0.05$], Talker Race x Vowel [$F(4, 396) = 8.3$, $p < 0.001$], and Picture x Gender [$F(1, 99) = 4$, $p < 0.05$]. The three-way interaction between Picture x Vowel x Gender was also significant [$F(4, 396) = 5$, $p < 0.001$].

Figure 1 shows the effect of selective imitation for low vowels. In this figure, the difference in distance measure on the *y*-axis indicates the amount of phonetic imitation. A negative value indicates phonetic imitation, zero indicates no change, and a positive value signals vocalic divergence. Blocks 4, 5, and 6 are Shadowing Blocks while Block 7 is the Post-task Block where auditory exposure to the model talker had ceased. Post-hoc Tukey tests indicate that both /ɑ/ and /æ/ are imitated more than /i o u/ ($p < 0.001$). In addition, participants accommodated toward /æ/ more than /ɑ/ ($p < 0.05$). Auditory exposure to the model talker modified the production of the low vowels, but the production targets of the other three vowels were unaffected by the model talkers' auditory targets. The effect of Block is also visible in Figure 1. Post-hoc tests found imitation to be cumulative across shadowing blocks. There was more imitation in Block 5 than 4 ($p < 0.05$) and more in Block 6 than 4 ($p < 0.001$). During auditory exposure (Blocks 4-6) in each pair-wise comparison, there was more accommodation than during the Post-task reading ($p <$ 0..001). Selective vowel imitation is responsible for the Vowel by Block interaction. Since only /æ/ and /ɑ/ were imitated, these vowels are responsible for the cumulative block effect. There were no differences in /i o u/ across the four blocks.

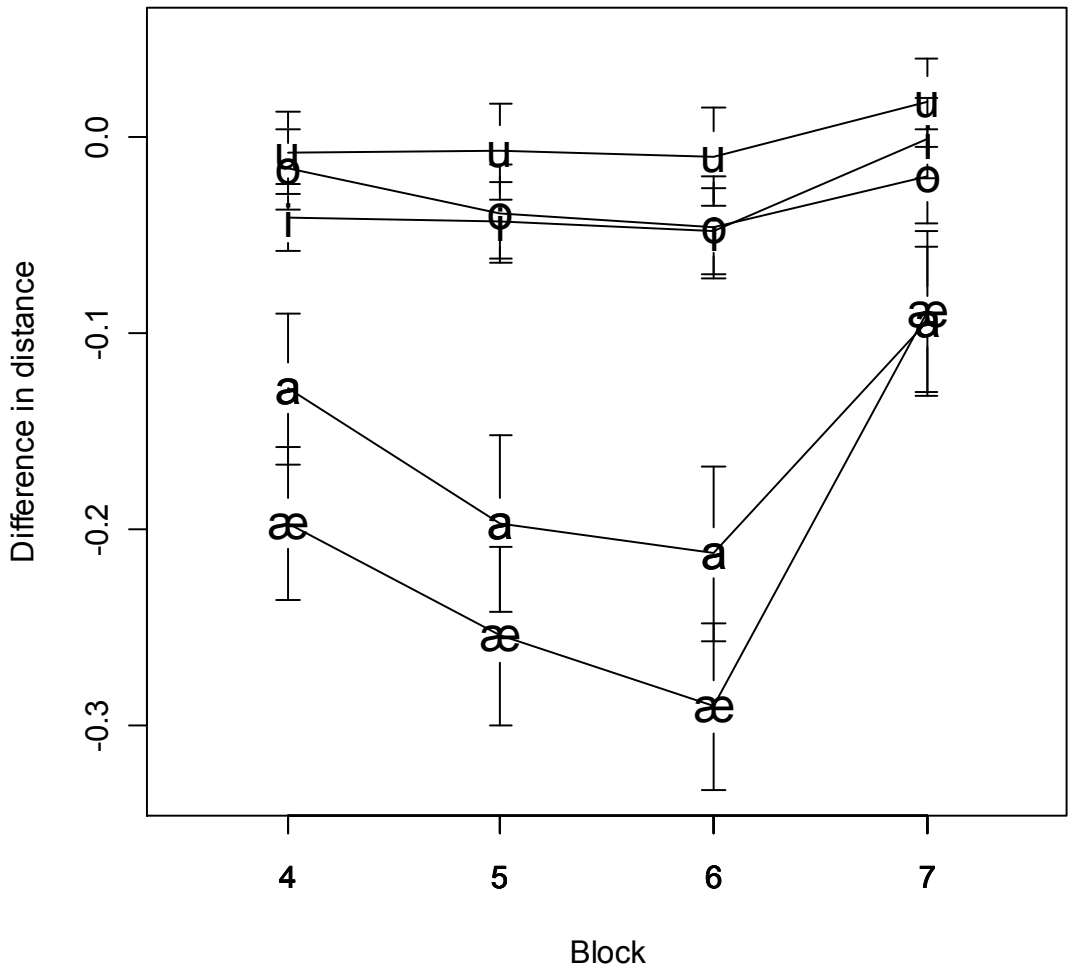

Figure 1. Spontaneous phonetic imitation for all participants by Vowel and Block. The Difference in Distance measure on the *y*-axis indicates the amount of phonetic imitation. A value of zero shows no change in vowel production as a result of auditory exposure to the model talkers. A negative value demonstrates phonetic imitation and a positive value demonstrates vocalic divergence. Blocks 4, 5, and 6 are Shadowing Blocks while Block 7 is the Post-task Block. The error bars represent 95% confidence intervals.

Figure 1 depicts the selective nature of vowel imitation, but based on the distance measurement made, information about the direction of the imitation is not available. Normalized formant plots from male and female participants are shown in Figures 2 and

3 for the Black talker and the White talker conditions, respectively. These figures show that the majority of vowel imitation comes from changes within the F1 dimension.

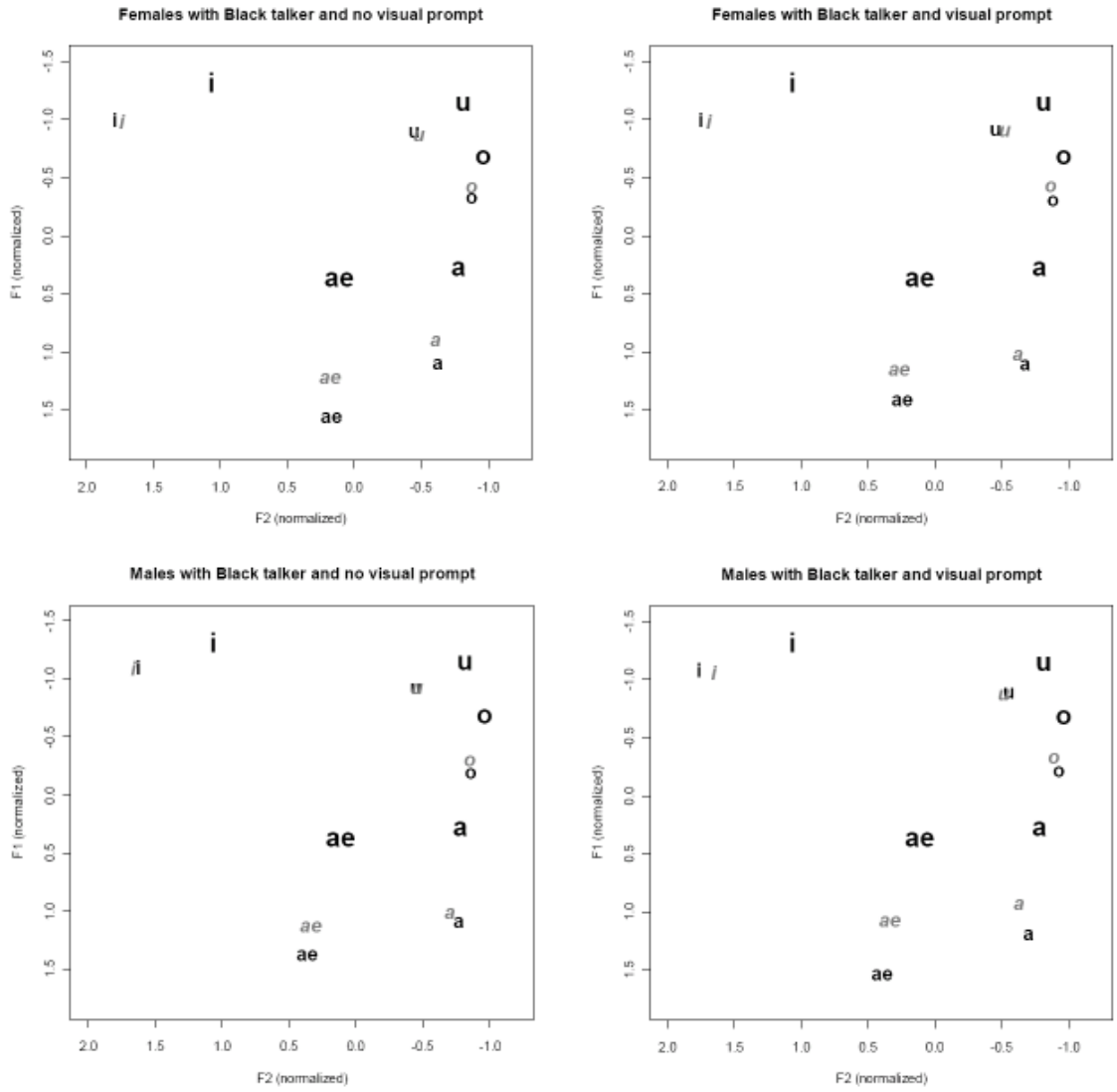

Figure 2. Formant plot displaying the direction of spontaneous phonetic imitation for the Black talker conditions. Normalized formant values are plotted. The mean of the model talkers' vowels are in the slightly larger font and in black. Participants' pre-task (Block 2) vowel means are plotted in a smaller black font and their productions from the final shadowing block (Block 6) are in the small italicized gray print. For the low vowels the shadowing productions move in the direction of the targets provided by the model talker.

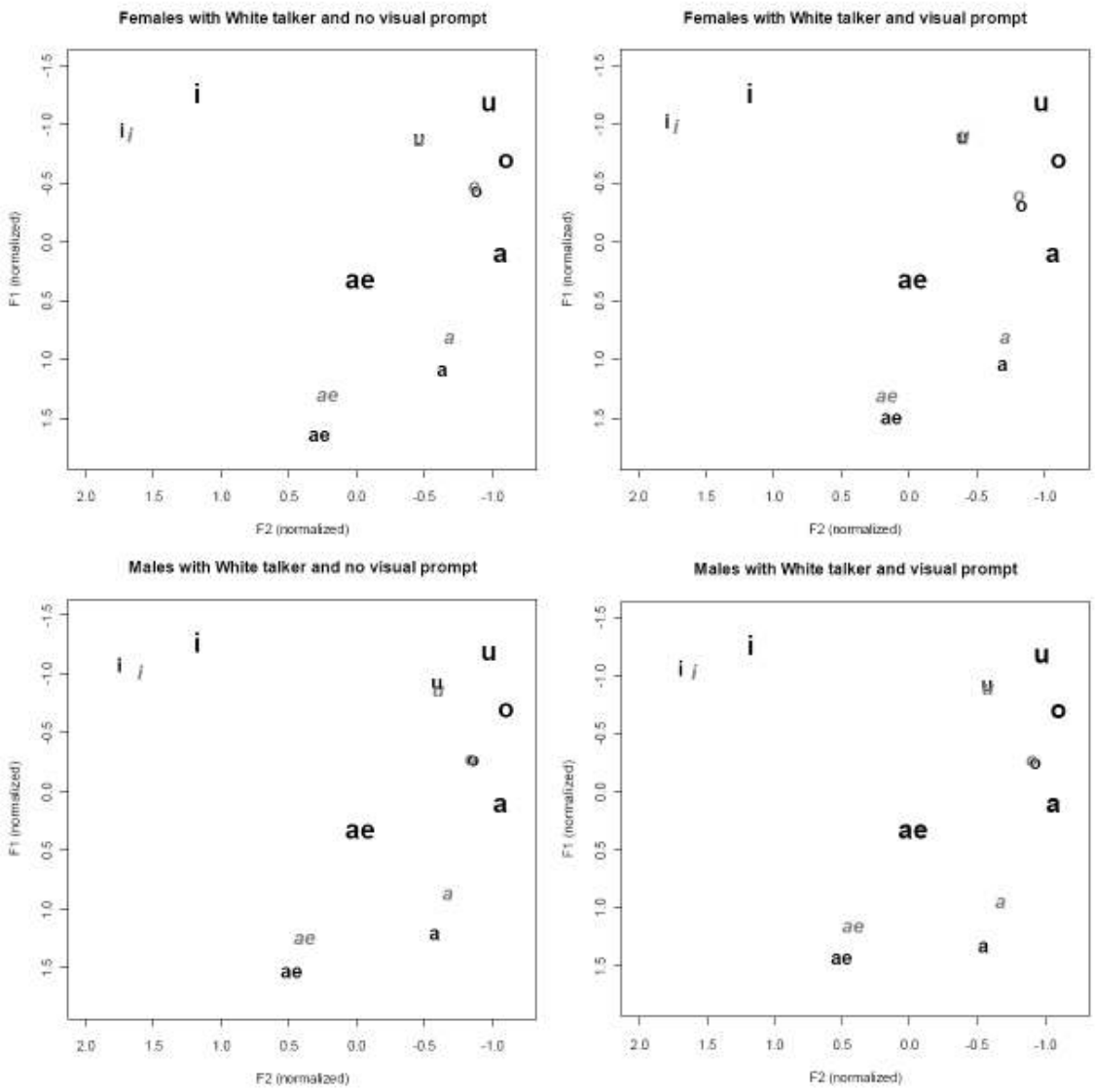

Figure 3. Formant plot displaying the direction of spontaneous phonetic imitation for the White talker conditions. Normalized formant values are plotted. The mean of the model talkers' vowels are in the slightly larger font and in black. Participants' pre-task (Block 2) vowel means are plotted in a smaller black font and their productions from the final shadowing block (Block 6) are in the small italicized gray print. For the low vowels the shadowing productions move in the direction of the targets provided by the model talker.

Males and females responded differently to the Visual Prompt and No Visual Prompt conditions. Difference in distance values for the these two conditions for male and females are summarized in Table 3. Male and female participants did not differ in terms of imitation in the No Visual Prompt condition [$t(1031) = 0.9$, p = n.s.], but male

participants imitated more than women in the Visual Prompt condition [$t(932) = -4.8$, $p = 0.001$]. In addition to this difference between genders in the Visual Prompt condition, there were differences within genders across the two conditions. Female participants accommodated to the vowels of the model talkers less in the Visual Prompt condition than in the No Visual Prompt condition [$t(1150) = -2.2$, $p < 0.05$] while male participants imitated vowels more in the Visual Prompt condition than they did in the No Visual Prompt condition [$t(954) = 3.6$, $p < 0.001$].

	Female	Male
No visual prompt	-0.09 (0.32)	-0.8 (0.31)
Visual prompt	-0.07 (0.31)	-0.13 (0.36)

Table 3. Mean *difference in distance* values for Females and Males in the Visual Prompt and No Visual Prompt Conditions. Standard deviations are provided in parentheses. A lower difference in distance value indicates more imitation. The genders did not differ from one another in the No Visual Prompt condition, but male participants imitated more in the Visual Prompt condition. The comparisons within gender are also significant; females imitated more in the No Visual Prompt condition than in the Visual Prompt condition while males imitated more in the Visual Prompt condition than in the No Visual Prompt condition.

The interactions between Talker Race x Vowel and Vowel x Block x Talker Race can be seen in Figure 4. This figure only shows the low vowels since only the low vowels elicited imitation in the task. Productions from the Black Talker condition are presented with a solid line and those from the White Talker condition are shown with a dashed line. Participants accommodated to /ɑ/ and /æ/ in the White Talker condition and /æ/ from the Black Talker condition to the roughly the same extent. However, the Black Talker /ɑ/ did not receive the same degree of imitation; /ɑ/ from the Black Talker

condition was still imitated more than /i/ ($p < 0.001$) and /u/ ($p < 0.001$). The paired comparison between /ɑ/ and /o/ in the Black Talker condition was just beyond significance ($p < 0.1$).

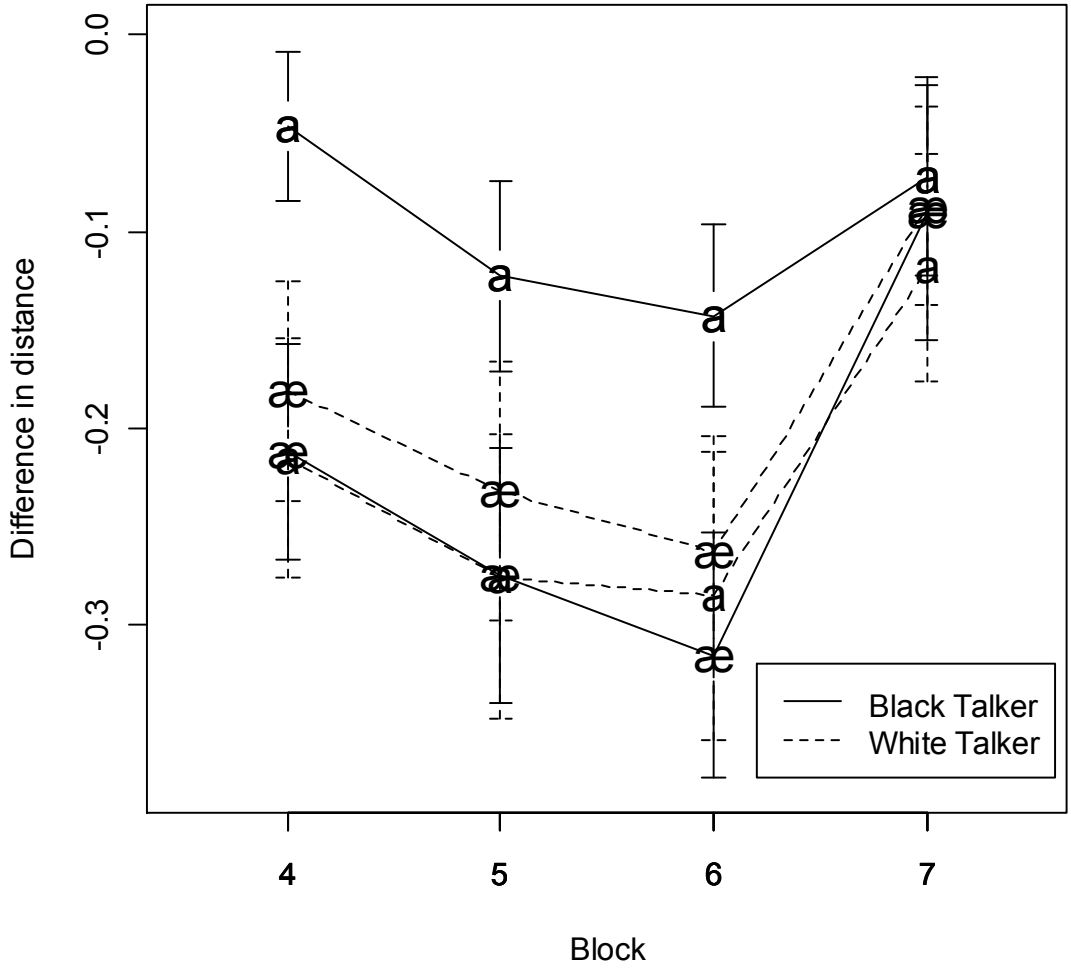

Figure 4. Spontaneous phonetic imitation for all participants for /ɑ/ and /æ/ by Talker Race. The Difference in Distance measure on the *y*-axis indicates the amount of phonetic imitation. A value of zero shows no change in vowel production as a result of auditory exposure to the model talkers. A negative value demonstrates phonetic imitation and a positive value demonstrates vocalic divergence. Blocks 4, 5, and 6 are Shadowing Blocks while Block 7 is the Post-task Block. Productions in the Black Talker condition have a

solid line and those in the White Talker Condition are dashed. The error bars represent 95% confidence intervals.

IV. DISCUSSION

To summarize the findings reported above, vowel imitation was found to be selective. This result is predicted by exemplar-based models of speech production that posit the activation of traces in speech perception lead to modifications in speech production. As predicted in the introduction, not all vowels changed as a result of auditory exposure to the model talkers. The vowels /ɑ/ and /æ/ elicited imitation that was cumulative across the shadowing blocks; more auditory exposure to the model talkers resulted in more imitation for the low vowels. Participant productions for the vowels /i o u/ went unchanged throughout the course of the task. Vowel accommodation for /ɑ/ and /æ/ resulted in shifts in the first formant frequency.

A. Preference for low vowels

There was a clear preference for imitating low vowels within the F1 dimension. During auditory exposure to talkers who produced low vowels with lower F1s, participants produced vowels with lower first formant frequencies. These results fall in line with theories of word-specific phonetics (Pierrehumbert, 2002). Talkers appear to stay within their pre-existing word-specific phonetic repertoires when producing words in this experimental paradigm. That is, participants may be selecting from pre-existing variants in order to approximate the auditory targets of the model talker, but they are not encoding new speech production targets simply for this task.

To understand how this view is compatible with the selective vowel results, consider the fact that there are spectral differences between prosodically accented and unaccented vowels (Lindblom, 1963; Engstrand, 1988). For example, in the sentence *The man IS going to the store*, as a content word *man* is stressed, but the prosodic focus of the sentence is on *is*, leaving *man* unaccented. However, in the sentence *The MAN is going to the store*, the subject *man* is both accented and stressed. The amount of formant expansion associated with accented and unaccented vowels is related to jaw movement, particularly within F1 (Summers, 1987; de Jong, 1995). Summers found greater differences in F1 for /æ/ than /ɑ/, suggesting that this vowel would have the largest array of production variants. This fact corresponds well with the current finding where /æ/ is the most imitated vowel. The tendency for the spectral differences between accented and unaccented vowels has been argued to be greater for low vowels overall, perhaps in relation to vowel sonority (Beckman et al., 1992).

B. Social effects

It has been argued previously that phonetic convergence is a socially mediated process (Pardo, 2006). In social psychology there is an entire framework – Communication Accommodation Theory (CAT; Giles, 1973; Giles and Coupland, 1991; Shepard et al., 2001) – that views imitation and convergent speech behavior as a social act that talkers use to modulate social distances in communication. In the experiment reported here there were several condition-specific results that can be considered further evidence that social factors mediate the extent of imitation. In the data reported above, male and female participants responded differently to the Visual Prompt and No Visual Prompt Conditions. Male participants imitated more in the presence of a picture; in the

presence of explicit social information about the identity of the talker, male participants were more likely to accommodate. On the other hand, female participants imitated less in the Visual Prompt condition. There were no differences in the extent of imitation between genders in the No Visual Prompt condition. Dijksterhuis and Bargh (2001) suggest that all types of behavioral information are automatic, but that social knowledge can inhibit or facilitate the imitation process. In the case of spontaneous phonetic imitation, social knowledge about the model talkers in the Visual Prompt condition inhibited accommodation by the females, but facilitated accommodation by the males. As a baseline measure and in the absence of social information, in the No Visual Prompt condition there were no differences between genders in terms of how much talkers imitate.

While there were no overall differences in how much the two talkers were imitated, participants accommodated toward the Black talker's /ɑ/ less than that of the White talker and less than either talker's /æ/. Since this finding did not interact with whether an image of the talker was presented, it is not likely that the lack of imitation involves the talker's race. Rather, there was something inherent in the Black talker's /ɑ/ that made it less likely to be imitated, although it still elicited more accommodation than /i o u/. While the vowel spaces of the two model talkers were comparable, as shown in Table 4, the Black talker's mean second formant for /ɑ/ is considerably higher than that of the White talker. Low vowels with lower F2 are the novel variants in most dialects of American English (Hinton et al., 1987; Moonwomon, 1991; Munson et al., 2006). The more novel productions from the White talker incited more vocalic imitation. Although it

should be noted that despite the novelty of the backed variant, imitation was still in the direction of F1. It is possible that the conservative /ɑ/ variant used by the Black talker inhibited participants desire to accommodate toward the vowel.

		/æ/	/ɑ/
Black Talker	F1	606	586
	F2	1609	1125
White Talker	F1	609	559
	F2	1533	939

Table 4. Average first and second formant frequencies for /æ/ and /ɑ/ for the model talkers in the task.

V. Conclusion

The present data bear on current discussion in the literature regarding the mechanisms at work in spontaneous phonetic imitation is in order. Exemplar-based models of speech production and perception are able to account for these results, if an adaptable model that allows for attention weighting is implemented (eg. Johnson, 1997; Goldinger and Azuma, 2003). The lack of equal imitation in all conditions indicates that a simple automatic exemplar-based model is untenable. In addition, since imitation is not a consistent consequence of auditory exposure, it is unlikely that production and perception work out of the same stores of exemplars. Although, the fact that auditory exposure can shift production targets at all does imply a strong connection between the two processes.

A highly important aspect of finding vowel accommodation in spontaneous phonetic imitation is that it demonstrates the labile nature of linguistic segments with respect to both their perceptual encoding and their variation in production. First, listeners

must perceive the detailed acoustic structure of an utterance in order to have those details influence their production. Second, in speech production, participants alter the characteristics of the output without modifying the categorical identity of the segment they produce. The exact selection of a production variant is influenced by auditory exposure. Lastly, this result demonstrates that part of the variability in speech production can be accounted for by auditory exposure during or prior to production.

Acknowledgments

This research was supported by the Center for Race and Gender and the Abigail Hodgen Publication Fund at the University of California, Berkeley. This project was conducted in partial fulfillment of the requirements of Doctor of Philosophy to the author at UC Berkeley. Thanks to Keith Johnson, Andrew Garrett, Rudy Mendoza-Denton, and Ben Munson for improving this project immeasurably. Special thanks to Dasha Bulatov and Tyler Frawley for help completing the project.

References

Adank, P., Smits, R., and van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. J. Acoust. Soc. Am. 116, 99-107.

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

Beckman, M.E., Edwards, J., and Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In *Papers in Lab. Phon. II*, ed. G. Docherty and D.R. Ladd, 68–86. Cambridge University Press.

Boersma, P., and Weenink, D. (2005). Praat: doing phonetics by computer (version 4.3.29) [computer program], from http://www.praat.org/. Institute of Phonetic Sciences, Amsterdam.

Chambers, J. (1992). Dialect acquisition. Language. 68:673–705.

Coupland, N. (2008). The delicate constitution of identity in face-to-face accommodation: A response to Trudgill. Lang. in Soc. 37:267–270.

Delvaux, V. and Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. Phonetica 64:145–173.

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

Engstrand, O. (1988). Articulatory correlates of stress and speaking rate in Swedish VCV utterances. J. Acoust. Soc. Am. 83.

Evans, B.G. and Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. J. Acoust. Soc. Am. 121:3814–3826.

Giles, H. (1973). Accent mobility: a model and some data. Anth. Ling. 15:87–105.

Giles, H. and Coupland, N. (1991). Language: Contexts and consequences. Milton Keynes: Open University Press.

Goldinger, S. D. (1997). Perception and production in an episodic lexicon. In *Talker Variability in Speech Processing*, (eds.) K. Johnson and J.W. Mullennix, 33–66. Academic Press: San Diego.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. Psych. Rev. 105:251–279.

Goldinger, S. D. and Azuma, T. (2003). Puzzle-solving science: the quixotic quest for units in speech perception. J. Phonetics 31:305–320.

Goldinger, S. D. and Azuma, T. (2004). Episodic memory in printed word naming. Psych. Bul. Rev. 11:716–722.

Goldstein, L. and Fowler, C.A. (2003). Articulatory phonology: A phonology for public language use. In (eds.) N. Schiller and A. Meyer *Phonetics and phonology in language comprehension and production: Differences and similarities*. 159-207. Berlin: Mouton de Gruyter.

Gregory, S.W., Dagan, K., and Webster, S. (1997). Evaluating the relation between vocal accommodation in conversational partners' fundamental frequencies to perceptions of communication quality. J. Nonverbal Behav. 21:23–43.

Gregory, S. W., Green, B.E., Carrothers, R.M., and Dagan, K.A. (2001). Verifying the primacy of voice fundamental frequency in social status accommodation. Lang. Comm. 21:37–60.

Gregory, S.W. and Hoyt, B.R. (1982). Conversation partner mutual adaptation as demonstrated by Fourier series analysis. J. of Psych. Res. 11:35–46.

Gregory, S.W. and Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. J. Person. Soc. Psych. 70:1231–1240.

Gregory, S. W., Webster, S., and Huang, G. (1993). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. Lang. Comm. 13:195–217.

Hinton, L., Moonwoman, B., Bremmer, S., Luthin, H., van Clay, M., Learner, J., and Corcoran, H. (1987). It's not just the Valley Girls: A study of California English. In Proc. 13th Ann. Meeting Berk. Ling. Soc., (eds.) J. Aske, N. Beery, L. Michaelis, and H. Filip, 117–28. Berkeley, CA: BLS.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In *Talker Variability in Speech Processing*, (eds.) K. Johnson and J.W. Mullennix, 145–165.

de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. J. Acoust. Soc. Am. 97(1):491-504.

Kraljic, T., Brennan, S.E., and Samuel, A.G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. Cognition 107:54–81.

Kraljic, T. and Samuel, A.G. (2005). Perceptual learning for speech: Is there a return to normal? Cog. Psych. 51:141–178.

Kraljic, T. and Samuel, A.G. (2006). Generalization in perceptual learning for speech. Psychonomic Bul. Rev. 13:262–268.

Kraljic, T. and Samuel, A.G. (2007). Perceptual adjustments to multiple speakers. J. Memory Lang. 56:1–15.

Kraljic, T., Samuel, A.G., and Brennan, S.E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. Psych. Sci. 19:332–338.

Kuhl, P. and Metzloff, A. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. J. Acoust. Soc. Am. 100:2425–2438.

Lindblom, B. (1963). Spectrographic study of vowel reduction. J. Acoust. Soc. Am. 35:1773–1781.

Lobanov, B. (1971) Classification of Russian vowels spoken by different listeners. J. Acoust. Soc. Am. 49, 606-08.

Maye, J., Aslin, R.N., and Tanenhaus, M.K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. Cog. Sci. 32:543–562.

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

Moonwomon, B. (1991). Sound change in San Francisco English. Ph.D. Dissertation, UC Berkeley.

Munro, M.J., Derwing, T.M., and Flege, J.E. (1999). Canadians in Alabama: a perceptual study of dialect acquisition in adults. J. Phonetics 27:385–403.

Munson, B., McDonald, E.C., DeBoe, N.L., and White, A.R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. J. Phonetics 34:202–240.

Namy, L.L., Nygaard, L.C., and Sauerteig, D. (2002). Gender differences in vocal accommodation: the role of perception. J. Lang. Soc. Psych. 21:422–432.

Natale, M. (1975a). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. J. Pers. Soc. Psych. 32:790–804.

Natale, M. (1975b). Social desirability as related to convergence of temporal speech patterns. Percept. Motor Skills 40:827–830.

Nielsen, K. (2008). The specificity of allophonic variability and its implications for accounts of speech perception. Ph.D. Dissertation, UCLA.

Norris, D., McQueen, J.M., and Cutler, A. (2003). Perceptual learning in speech. Cog. Psych. 47:204–238.

Nye, P.W. and Fowler, C.A. (2003). Shadowing latency and imitation: the effect of familiarity with the phonetic patterning of English. J. Phonetics 31:63–79.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. J. Acoust. Soc. Am. 119:2382–2393.

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

Payne, A. C. (1980). Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. In *Locating language in time and space*, (ed.)W. Labov, 179–218. Academic Press: New York.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In *Frequency effects and the emergence of lexical structure*, (eds.) J. Bybee and P. Hopper, 137–157. Amsterdam: John Benjamins.

Pierrehumbert, J. (2002). Word-specific phonetics. In Papers in Lab. Phon. VII , eds. C. Gussenhoven and N. Warner, 101–139. Berlin: Mouton de Gruyter.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. Lang. and Speech 46:115–154.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). E-Prime: User's Guide, version 1.0. Psychology Software Tools.

Shepard, C. A., Giles, H. and LePoire, B.A. (2001). Communication accommodation theory. In *The New Handbook of Lang. and Social Psychology*, (eds.) W. P. Robinson and H. Giles, 33–56. John Wiley & Sons Ltd.

Shockley, K., Sabadini, L., and Fowler, C.A. (2004). Imitation in shadowing words. Percept. Psychophys. 66:422–429.

Summers, W. V. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. J. Acoust. Soc. Am. 82:847– 863.

Trudgill, P. (1986). Dialects in contact. Blackwell Publishing.

Trudgill, P. (2008). Colonial dialect contact in the history of European Languages: On the irrelevance of identity in new-dialect formation. Lang. in Soc. 37:241–280.