Journal of Phonetics 40 (2012) 177–189

Contents lists available at SciVerse ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Evidence for phonetic and social selectivity in spontaneous phonetic imitation

Molly Babel *

Department of Linguistics, University of British Columbia, Totem Fields Studios, 2613 West Mall, Vancouver, British Columbia, Canada V6T 1Z4

ARTICLE INFO

Article history:
Received 28 May 2010
Received in revised form
14 July 2011
Accepted 2 September 2011
Available online 27 October 2011

ABSTRACT

Spontaneous phonetic imitation is the process by which a talker comes to be more similar-sounding to a model talker as the result of exposure. The current experiment investigates this phenomenon, examining whether vowel spectra are automatically imitated in a lexical shadowing task and how social liking affects imitation. Participants were assigned to either a Black talker or White talker; within this talker manipulation, participants were either put into a condition with a digital image of their assigned model talker or one without an image. Liking was measured through attractiveness rating. Participants accommodated toward vowels selectively; the low vowels /æ ɑ/ showed the strongest effects of imitation compared to the vowels /i o u/, but the degree of this trend varied across conditions. In addition to these findings of phonetic selectivity, the degree to which these vowels were imitated was subtly affected by attractiveness ratings and this also interacted with the experimental condition. The results demonstrate the labile nature of linguistic segments with respect to both their perceptual encoding and their variation in production.

1. Introduction

Language users acquire the varieties and dialects spoken around them. Sentence structure, lexical selection, and pronunciation are all determined by patterns in the ambient language. Children, for example, acquire language from their caretakers and peer group (Chambers, 1992; Payne, 1980). When adults move to new dialect areas the adoption of novel features is more variable, but, generally, they eventually acquire aspects of the new dialect (Evans & Iverson, 2007; Munro, Derwing, & Flege, 1999; Trudgill, 1986), even after brief exposure (Delvaux & Soquet, 2007). In the absence of physical relocation, adults' speech patterns are also susceptible to change (Harrington, 2006, 2007; Harrington, Palethrope, & Watson, 2000a,b). These new dialect features do not wholly replace the native dialect, but rather are added to a speaker's repertoire (Howell, Barry, & Vinson, 2006). This addition, in theory, would enable a talker to shift dialects or styles. A question surrounding such style-shifting behaviour is whether it is within a speaker's conscious control (cf. Eckert, 2001: 124; Labov, 2001: 85).

In addition to changing production as a result of real-world exposure to a new dialect, similar changes in speech have been induced in the laboratory. For example, imagine a picture with a man holding a cake and a woman standing by. The orientation of the image suggests the man intends to pass the cake to the woman. Participants who have been exposed to an oral description *The boy gave the toy to the teacher* are more likely to describe the cake picture as *The man gave the cake to the woman* as opposed to *The man gave the woman the cake*. The second description is a completely grammatical utterance that accurately conveys what is going on in the picture. But, simply being exposed to the construction *give X to Y* favours the future use of that construction over *give Y X* (Bock, 1986; Branigan, Pickering, McLean, & Cleland, 2007; Pickering & Ferreira, 2008). Similarly, with word choice, individuals align their lexical selections to those of their interlocutors in experimental settings (Garrod & Doherty, 1994).

These constant updates to the language system are particularly interesting from a phonetic perspective. The fact that the perceptual categories of language are labile is well grounded (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Kraljic & Samuel, 2005, 2006, 2007; Kraljic, Brennan, & Samuel, 2008; Kraljic, Samuel, & Brennan, 2008; Maye, Aslin, & Tanenhaus, 2008; Norris, McQueen, & Cutler, 2003). With respect to speech production, when faced with the question why do talkers sound like they do, the immediate answer is that it is determined by a talker's physiology. The exact size and shape of talker's oral cavity go a long way in determining the acoustic characteristics of that particular talker. Putting aside physiological differences between talkers, however, the question of why we sound like we do returns us to the notion of sounding like those around us. From birth, talkers acquire the ambient language and dialect to which they are exposed, and, if we are to examine how

* Tel.: +1 604 822 6322; fax: +1 604 822 9687.
E-mail address: molly.babel@ubc.ca

0095-4470/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2011.09.001

these speech production targets may change from one production to the next, we simply need to examine phonetic imitation.

Phonetic imitation, also known as phonetic convergence or phonetic accommodation, is the process in which a talker takes on acoustic characteristics of the individual they are interacting with. Like sentence structure and lexical selection, phonetic imitation has also been investigated in the laboratory. This research has revealed that phonetic imitation can occur in both socially minimal situations where talkers are simply producing single words (Goldinger, 1997, 1998; Goldinger & Azuma, 2004; Namy, Nygaard, & Sauerteig, 2002) and in cooperative, socially rich, dyadic interactions (Pardo, 2006, 2009). This phenomenon is important, since it may account for a wide range of phenomena such as historical sound change and dialect acquisition (Trudgill, 2008). Moreover, the cognitive mechanisms that motivate spontaneous imitation are pervasive from infancy (Kuhl & Metzloff, 1996) and existent across non-linguistic behaviours like foot-shaking and face-touching (Dijksterhuis & 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. The topic is particularly crucial for theories that propose a strong link between the two processes (e.g., Goldstein & Fowler, 2003) as well as for theories about the representational units that mediate perception–production link.

For example, with respect to representation, Mitterer and Ernestus (2008) argue that abstract phonological units mediate the link between speech perception and speech production. Their argument is based on evidence gathered through a speeded shadowing task where they demonstrate that Dutch speakers do not accommodate to different speakers' /r/ variants or their degree of pre-voicing. Using a slightly modified version of this paradigm where the task is, crucially, not speeded, other researchers have shown that speakers phonetically imitate model speakers' productions (Goldinger, 1998; Namy et al., 2002). Investigations of the details involved in phonetic imitation reveal that that speakers imitate subphonemic properties of the speech signal (Nielsen, 2008, 2011; Shockley, Sabadini, & Fowler, 2004), which robustly suggests that a more detailed phonetic signal bridges the relationship between perception and production in speech. The current experiment is an auditory naming task that uses methods similar to previous work (Goldinger, 1998; Namy et al., 2002; Shockley et al., 2004), but focuses on the phonetic details of vowel imitation. Using two model talkers across four conditions, the results support the claim that subphonemic phonetic details are mapped from speech perception and production in ways that are mediated by experimental condition and talker voice. Crucially, these data also suggest a relationship between social liking and the strength of the mapping. This finding indicates that the socially mediated perception-behaviour link proposed by Dijksterhuis and Bargh (2001) applies to speech behaviour as well.

1.1. Phonetic imitation

The earliest studies on imitation and convergence examined broad acoustic measures and relied 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 behaviours 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 & Hoyt, 1982) and long-term average spectra (Gregory, Webster, & Huang, 1993; Gregory & Webster, 1996; Gregory, Dagan, & Webster, 1997; Gregory, Green, Carrothers, & Dagan, 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) interacted 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 behaviour. 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 in finding that female participants accommodated to the model talkers more than male participants.

One of the most influential pieces of 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 (2009) provides acoustic analyses 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. Pardo's analysis also involved comparing formant frequencies of /i u æ ɑ/ from a pre- and post-task *hVt* wordlist. The analysed items were words not produced as part of the map task that comprised the core of Pardo's analysis, so any changes in the wordlist items would be indicative of persistent system-wide changes. Pardo found that givers of map instructions centralized their low vowels from pre- to post-task, which also involved diverging from the receivers to whom they were giving directions.

Thus far, research has established that talkers do spontaneously imitate and accommodate in verbal interaction. Still, little is known about what within the acoustic signal may be the target of imitation. 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 (2011) 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). Analyses revealed a word-specificity effect such that increased VOT effects were strongest in the /p/-initial words heard in the exposure phase. Sancier and Fowler (1997) examined how VOT patterns transfer cross-linguistically based on the ambient language environment for a bilingual speaker of Brazilian Portuguese and English. They found that after multi-month spans of speaking English (her L2), the VOT of voiceless stops increased in her Brazilian Portuguese productions. The reverse was also observed, such that her English voiceless stop VOTs decreased in duration after being in a Brazilian Portuguese environment. Turning to vowels, a study by Vallabha and Tuller (2004) examined talkers' abilities to intentionally imitate their own steady-state vowel productions. They found that their three participants exhibited talker-specific patterns of imitation bias, which could either be due to noise in production or bias in the perceptual processing of the tokens. Vallabha and Tuller speculate that the dialect background of their participants could also have influenced the directions of the imitation bias.

While models of speech perception and production may account for imitation through the structure of the cognitive language system (Pickering & Garrod, 2004), others have accounted for accommodation in speech in its function as a social tool. Communication Accommodation Theory (CAT) argues that speech convergence phenomena are motivated by an individual's motivation to be socially accepted or identify with a particular social group (Giles & Coupland, 1991: 71–72). Shepard, Giles, & LePoire (2001: 33–34) posit that social and cognitive processes affect linguistic behaviour and that individual motivation is the driving force behind accommodative speech behaviour. Language is, therefore, a tool used by speakers to achieve a particular social distance. For example, talkers converge with an interlocutor in order to lessen the social distance and they diverge to increase social distance (Bourhis & Giles, 1977; Giles, 1973; Giles, Taylor, & Bourhis, 1973). Similarly, Bell's theory of audience design (Bell, 1984, 2001) argues that talkers will shift speech styles in accordance with audience – interlocutor, overhearers, eavesdroppers, etc. – speech styles, such that in a given instance, a talker's speech style reflects the demands of the setting.

1.2. The current experiment

This background leads to the goals of the current experiment, which are multifold. The first goal is to expand our understanding about *what* is imitated in the phonetic signal by targeting the investigation on the first and second formant frequencies of vowels. While exploring vowels, the phenomenon of phonetic imitation provides a nice test case for theories of speech perception and production. Theories of speech perception that hypothesize that speech gestures are the objects of speech perception such as the Motor Theory (Liberman, 1957; Liberman & Mattingly, 1985) and Direct Realism (Fowler, 1996; Goldstein & Fowler, 2003) easily account for imitation in speech as the perceived signal essentially contains the specific gestural instructions as to *how* to imitate. Exemplar-based models of speech perception (Goldinger, 1997; Johnson, 1997; Pierrehumbert, 2001, 2003) also easily account for imitative phenomena. Within an exemplar-based model episodic traces are activated in memory when a talker voice or word is perceived. Recently, Tilsen (2009) presented evidence for exemplars in speech production through a subphonemic shadowing task. 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. Such a theory predicts that upon perceiving a particular word, the activated traces will contribute to its production. Thus, in terms of phonetic imitation, exposure to words produced by a model talker will shift a participant's productions toward those of the model talker. In addition, given the assumptions of exemplar-based theories, the effect of imitation should be cumulative. That is, higher levels of activation, which are expected to result from repeated exposure, should lead to more imitation; this was essentially found by Goldinger (1998). In that study listeners judged more imitation for tokens that had higher repetition counts in the immediate shadowing conditions. A crucial difference between these two theories is that there exist formulations of exemplar-based models, which involve the incorporation of social cognition directly into the model of spoken language (e.g., Johnson, 2006).

The second and third goals of the experiment relate to the social nature of imitation and how it relates to levels of social engagement. If phonetic imitation is a primarily social behaviour, then it should occur more in situations that have more social context. In this experiment, one set of conditions includes Visual Prompts, which are still digital images of the model talkers. This manipulation is intended to increase the level of cognitive social engagement with the task by embodying the model voices with the talkers' faces. Thus, the second goal is to compare imitation across Visual Prompt and No Visual Prompt Conditions, where the presence of the social information available in the image is predicted to alter participants' behaviours (Hay, Warren, & Drager, 2006). The third goal is an attempt to incorporate social and behavioural psychology into our understanding of phonetic imitation. To review, exemplar-based theories predict imitation will occur naturally as a function of the linguistic system and social theories predict imitation will occur in response to more highly crafted social factors. Previous work on phonetic imitation demonstrates that it occurs in relatively asocial laboratory tasks (Goldinger, 1997, 1998; Namy et al., 2002; Nielsen, 2008, 2011; Shockley et al., 2004), which suggests that it is a low-level automated behaviour. Such findings, however, do not necessarily preclude that social factors cannot play a role and, of course, a speech research laboratory is still a type of social context. As mentioned above, Dijksterhuis and Bargh (2001) argue that the most simple mitigating factor in the perception-behavioural link is liking. From this we can predict that an individual's degree of liking of a model talker should affect their degree of imitation. In this experiment liking is measured through participant ratings of model talker attractiveness, which is highly correlated with liking (Bryne, 1971). Specifically, CAT offers the prediction for this study that when there is interpersonal liking between the participant and the model talker we can expect convergent speech behaviour. Cases of linguistic divergence are then predicted where the participant does not like the model talker and rates the model talker as unattractive.

2. Methodology

The methods for the auditory naming task are described below.

2.1. Stimuli

Fifty low frequency monosyllabic words from CELEX (Baayen, Piepenbrock, & van Rijn, 1993) with the vowels /i æ ɑ o u/ were selected as stimuli (see Table 1). Low frequency words were

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

/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

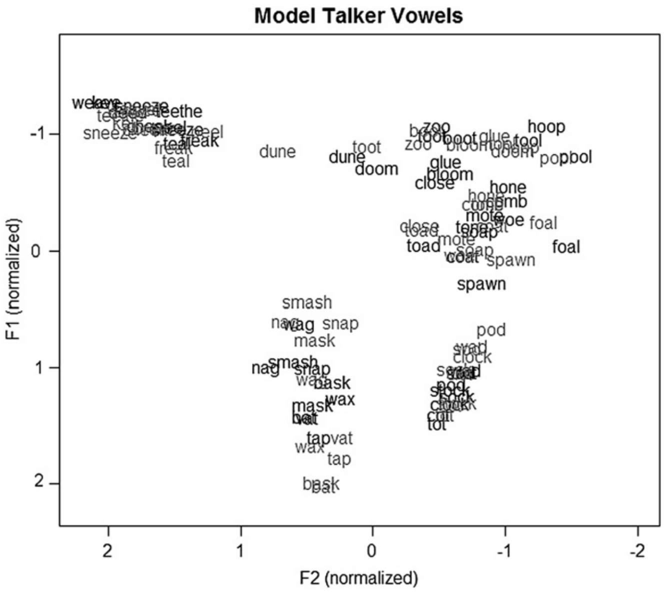

Fig. 1. Stimuli used in the task as produced by the model talkers. The tokens printed in black are those of the Black model talker and those in dark grey are from the White model talker.

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 any speech, language, or hearing disorder. They were both compensated $10 for their time.

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, Eschman, & Zuccolotto, 2002). Productions were digitally recorded to the hard drive of a PC at a 44.1 K sampling rate. 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 tokens used in the experiment are presented in Fig. 1 for both model talkers. The talkers were also digitally photographed from the mid-chest and up for the visual stimuli.

2.2. Participants

One hundred and eleven self-identified White participants (male$=$51) 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. The assignment of participants to the experimental conditions is reported in Table 2. All participants were compensated $10 for their time.

2.3. 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), and 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 in to the side of the mouth and AKG K240 headphones. Word productions were digitally recorded to the hard drive of a PC at a 44.1 K sampling rate. Recordings were down-sampled to 22 K before analysis.

Table 2
The four experimental conditions and their design along with the number of male and female participants assigned to each condition. 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 (or visual prompt) of the talker presented in the condition.

	Voice	Picture	Participants
Black talker with No Visual Prompt	Black	None	$m=13, f=15$
Black talker with Visual Prompt	Black	Black	$m=13, f=14$
White talker with No Visual Prompt	White	None	$m=11, f=14$
White talker with Visual Prompt	White	White	$m=12, f=15$

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 photo (sized 466×366 pixels) of the talker the participant was presented with was shown on the screen for the duration of the shadowing portion of the task. In this condition the Black talker's photo was always presented with the Black talker's voice and the White talker's photo was always presented with the White talker's voice. 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 race of the Black talker; both the Black and White talker were both identified as White by participants ($\chi(1)=0.08$, $p=$n.s.). Participants in the Visual Prompt Conditions were asked to rate the attractiveness of the model talker on a scale of 1–10, where 1 was the most unattractive rating and 10 was the most attractive. These attractiveness data are used in the analysis in Section 4.2.

3. Analysis

3.1. Auditory naming task

A Praat (Boersma & 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, Smits, and van Hout (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 distance value was calculated of how the production of a particular word evolves through the course of the experiment is needed. To this end, the Euclidean distance in the F1 $\times$ F2 space was calculated from each participant production to the model talker production of the same word from the talker to which they had been assigned; this was done using the formula: square root$((\mathrm{word}_{\mathrm{model's\ F1}}-\mathrm{word}_{\mathrm{participant's\ F1}})^2+(\mathrm{word}_{\mathrm{model's\ F2}}-\mathrm{word}_{\mathrm{participant's\ F2}})^2)$. 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. Note that since this difference in distance value includes the original baseline distance, the distance values from the pre-task block are not included in the analyses below. Fig. 2 presents histograms for female and male participants of each participants' averaged difference in distance value. The dashed vertical line denotes the 0 difference in distance point of no change. The majority of subjects fall on the negative side of the scale, indicating the tendency to accommodate to the vowel spectra of the model talker.

4. Results

4.1. Analysis across conditions

The histograms in Fig. 2 illustrate the peaks in the distributions are negative, suggesting an overall trend of imitation. In order to determine whether there were overall effects of imitation for the entire data set and for each vowel across all participants, a series of six *t*-tests were conducted to determine whether the difference in distance values were significantly below 0. A Bonferroni correction adjusted the significant alpha level for the six comparisons to $p=0.008$. Overall, difference in distance values were significantly below 0 [$M=-0.026$, SD$=0.27$; $t(36{,}374)=-17.99$, $p<0.001$]. For each vowel, difference in distance values were all also significantly below 0: /i/ [$M=-0.008$, SD$=0.19$; $t(7299)=-3.743$, $p<0.001$]; /æ/ [$M=-0.07$, SD$=0.34$; $t(7268)=-17.7$, $p<0.001$]; /ɑ/ [$M=-0.18$, SD$=0.31$; $t(7237)=-4.87$, $p<0.001$]; /o/ [$M=-0.012$, SD$=0.23$; $t(7162)=-4.3$, $p<0.001$]; /u/ [$M=-0.02$, SD$=0.26$; $t(7404)=-6.72$, $p<0.001$]. This set of findings establishes that globally, there were effects of imitation across the entire data set and for each of the vowels. The following analyses explore the differences in imitation and the patterns across groups that arose as a result of the experimental manipulations. First, the results of an ANOVA on the full design are reported. Subsequently, this is broken down into the No Visual Prompt and Visual Prompt Conditions for further analysis.

The full design of the experiment was a 2 (Voice: Black talker or White talker) $\times$ 2 (Picture: Visual Prompt or No Visual Prompt) $\times$ 5 (Vowel: i æ ɑ o u) $\times$ 4 (Block: Task Blocks 1, 2, 3 and the Post-task Block) $\times$ 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. There were main effects of Block [$F(3,\ 297)=23.09$, $p<0.001$] and Vowel [$F(4,\ 396)=7.61$, $p<0.001$]. These factors were also significant as a two-way Vowel $\times$ Block interaction [$F(12,\ 1188)=2.07$, $p<0.05$]. Vowel also interacted significantly with Gender [$F(4,\ 396)=3.4$, $p<0.01$]. There three-way interactions between Voice $\times$ Block $\times$ Vowel [$F(12,\ 1188)=2.03$, $p<0.05$] and Gender $\times$ Block $\times$ Vowel [$F(12,\ 1188)=2.64$, $p<0.01$]. A four-way interaction of Picture $\times$ Voice by Block by Vowel [$F(12,\ 1188)=2.2$, $p<0.01$] and a five-way interaction of Picture by Voice by Sex by Block by Vowel [$F(12,\ 1188)=2.98$, $p<0.001$] was also found.

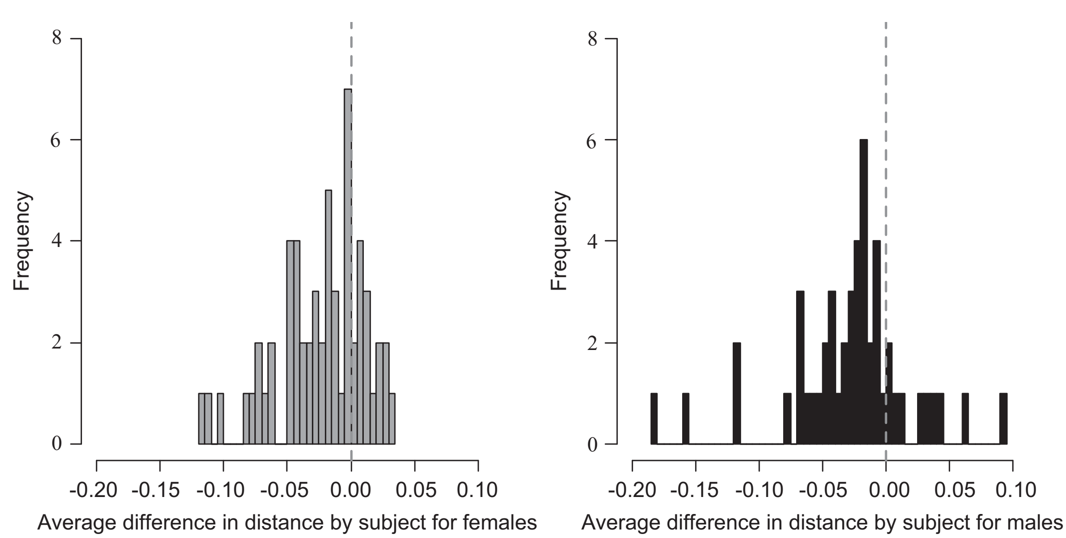

Fig. 2. A histogram of each participants' averaged difference in distance value. Female participants are on the left panel and male participants are on the right. The dashed vertical line denotes the 0 point of no change. A negative value indicates accommodation, while a positive value indicates divergence.

A main effect of the presence or absence of the model talker's picture was just beyond the level of significance [$F(1, 99)=3.38$, $p=0.69$]. A post-hoc t-test comparing difference in distance values in the Visual Prompt ($M=-0.031$, $SD=0.12$) and No Visual Prompt Conditions ($M=-0.016$, $SD=0.1$) revealed significantly more imitation in the Visual Prompt Condition [$t(2067)=3.21$, $p<0.01$]. Given this finding and the fact that Visual Prompt interacted significantly with other dependent measures and the overall complexity of the model, the data were subset into a Visual Prompt and No Visual Prompt group and submitted to separate analyses of variance.

Fig. 3 shows how Block and Vowel interact in imitation for the No Visual Prompt Conditions. This figure suggests that imitation occurred more strongly in the shadowing blocks; the effects of imitation are attenuated in the Post-task block in the No Visual Prompt Conditions. Both Figs. 3 and 4 illustrate vowel-specific patterns for the No Visual Prompt Conditions. Vowel is in interaction with Block in Fig. 3 and with participant gender in Fig. 4. Fig. 4 demonstrates how female participants were more likely to imitate /æ/ than male participants. To confirm these observations, the No Visual Prompt data were entered into an analysis of variance with the difference in distance values as the dependent measure and Voice, Gender, Block, and Vowel as dependent measures. The model returned main effects of Block [$F(3, 147)=14.82$, $p<0.001$] and Vowel [$F(4, 196)=7.97$, $p<0.001$] and significant two-way Gender × Vowel [$F(4, 196)=5.4$, $p<0.001$] and Block × Vowel [$F(12, 588)=3.4$, $p<0.001$] interactions. Tukey's Honest Significant Difference tests were used for post-hoc testing. These tests correct for multiple comparisons (Baayen, 2008) and since they follow from the F-statistics and degrees of freedom in their preceding ANOVAs, only p-values are produced.

The effect of this Vowel and Block interaction for the No Visual Prompt Condition is shown in Fig. 3. Tukey post-hoc testing revealed there was significantly more imitation in Blocks 1 and 2 compared to the Post-task Block ($p<0.001$ and $p<0.05$, respectively). The Gender × Vowel interaction is presented in Fig. 4. As observed previously, the Vowel effects were moderated by participant gender. Post-hoc Tukey tests confirmed the observation that females imitated /æ/ more than males ($p<0.001$) and males imitated

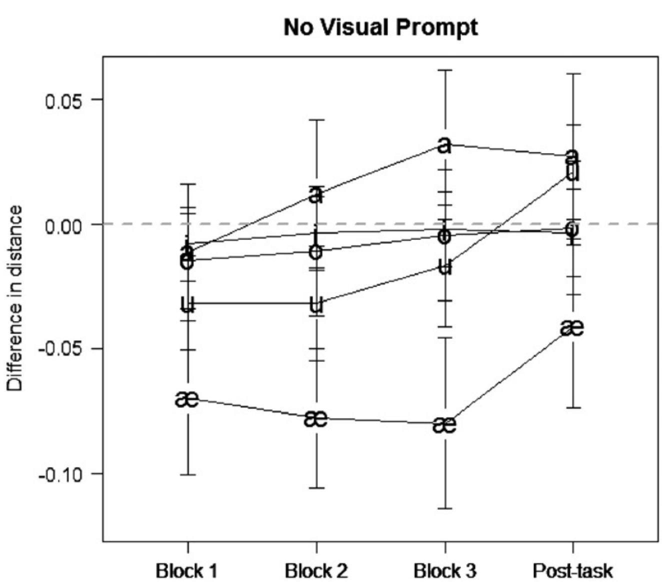

Fig. 3. Interaction of block and vowel in the No Visual Prompt Conditions. The difference in distance measure on the y-axis indicates the amount of phonetic imitation. A value of zero shows no change in vowel production as a result of auditory exposure to the model talkers. A negative value demonstrates phonetic imitation and a positive value demonstrates vocalic divergence. Task Blocks 1, 2, and 3 are shadowing blocks; the Post-task Block is labelled as such. The error bars represent 95% confidence intervals.

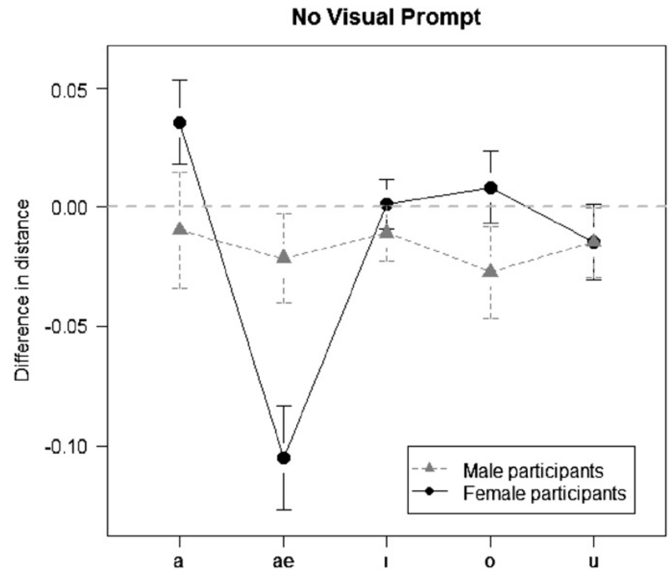

Fig. 4. Interaction of vowel and gender in the No Visual Prompt Conditions. The difference in distance measure on the y-axis indicates the amount of phonetic imitation. A value of zero shows no change in vowel production as a result of auditory exposure to the model talkers. A negative value demonstrates phonetic imitation and a positive value demonstrates vocalic divergence. The error bars represent 95% confidence intervals.

/ɑ/ more than females ($p<0.05$); there were no other differences across genders for the other vowels.

Fig. 5 shows similar data for the Visual Prompt Conditions separated into different panels for the four combinations of Gender and Talker Race. The top two panels are for participants in the Black talker Visual Prompt Condition and the bottom two panels are for the parallel design with the White talker. The left panels are for female participants, and those on the right are for the male participants. The patterns here are quite complex; there is a considerable amount of variation within and across conditions. Some of this variability is likely due to uncontrollable factors like participant inattentiveness to the experiment. Overall, however, with the Black talker, both males and females tend to imitate /æ/ more. With the White talker, females also exhibited more /æ/ imitation, while males imitated /ɑ/ more with the White talker.

To test these observations, the Visual Prompt Conditions were submitted to an analysis of variance with the difference in distance value as the dependent measure and Voice, Gender, Block, and Vowel as dependent measures. This model found main effects of Block [$F(3, 150)=0.03$, $p<0.001$] and Vowel [$F(4, 200)=3.02$, $p<0.05$]. There were significant three-way Voice × Block × Vowel [$F(12, 600)=3.24$, $p<0.001$] and Gender × Block × Vowel [$F(12, 600)=2.65$, $p<0.01$] interactions. The four-way interaction between Voice × Gender × Block × Vowel was also significant [$F(12, 600)=3.1$, $p<0.001$]. This array of complex interactions is shown across the four panels in Fig. 5. Post-hoc Tukey tests found no differences between blocks in the Visual Prompt Conditions, but did establish that overall /æ/ was imitated more than the other vowels ($p<0.001$), while /ɑ/ was imitated more than /i/ and /o/ ($p<0.05$). A planned t-test found male participants in the White talker Visual prompt Condition imitated /ɑ/ more than /æ/ [$t(94.6)=-2.26$, $p=0.02$], in contrast to /æ/ being the most imitated vowel in all other conditions; this effect is visible in the lower right panel of Fig. 5.

Thus far, the figures depict the general nature of vowel imitation, but based on the distance measurement made, information about the direction of the imitation is not available from these figures, nor

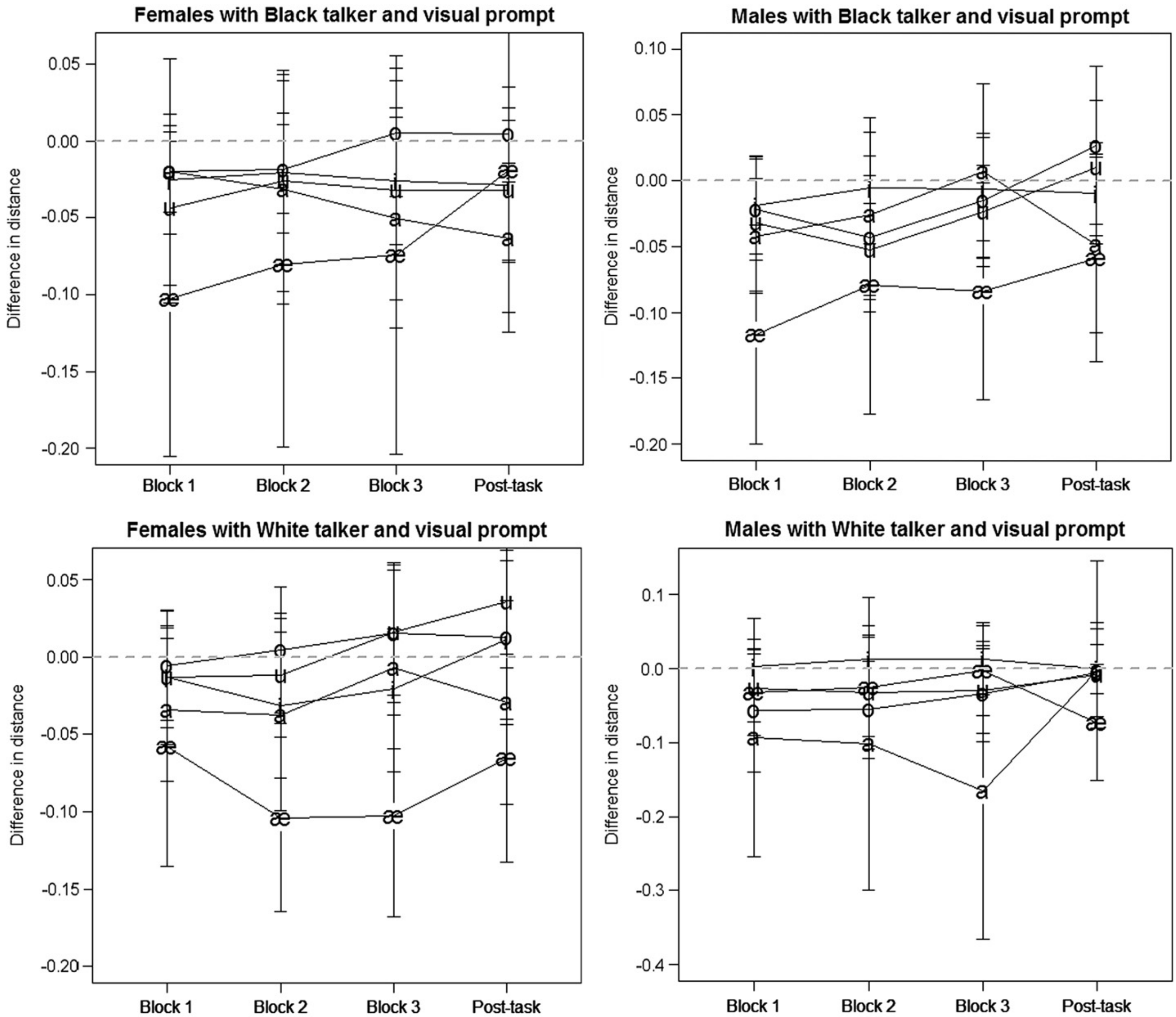

Fig. 5. Interaction of Block × Vowel × Gender × Talker in the Visual Prompt Conditions. The difference in distance measure on the *y*-axis indicates the amount of phonetic imitation. A value of zero shows no change in vowel production as a result of auditory exposure to the model talkers. A negative value demonstrates phonetic imitation and a positive value demonstrates vocalic divergence. Task Blocks 1, 2, and 3 are Shadowing Blocks; the Post-task Block is labelled as such. The error bars represent 95% confidence intervals.

from the analyses. Normalized formant plots from female and male participants are shown in Figs. 6 and 7 for all condition combinations. In these figures the average model talker vowels are plotted in the larger black font, for each condition participants' averaged baseline productions are in the smaller black font, and the averaged shadowed productions from the three shadowing blocks are in grey italics. As can be seen in these figures, the shadowed productions tend to move in the direction of the model talker's averaged token. For female participants, direct overlap is never obtained; note, however the male panel presenting the results from the White talker with No Visual Prompt Condition, where the shadowed /æ/ completely overlaps with the model talker's averaged token. These results are not in conflict with the prior statistical reports regarding the amount of phonetic distance travelled in the course of imitation. While female participants went a greater comparative distance and therefore imitated /æ/ *more* than males, the male participants all began the task with a smaller phonetic distance (even after normalizing for physiological differences) than the female participants, which results in overlapping with the model talker's tokens despite not having imitated more quantitatively than the female participants.[1]

[1] Much of the change in the vowels is in the direction of participants lowering their F1s for the low vowels. This raises the question as to whether the effects reported here are the result of the reduction. To examine whether it was the simple act of repeating the words was responsible for these patterns a group of 12 participants completed an experiment that used the same basic format as the shadowing task, but had listeners continue to read the words aloud in what were the shadowing blocks in the original design. That is, participants produced each word 8 times just as those in the shadowing task did. Participant productions were Lobanov normalized. Since there was no model talker to compare difference in distance to, the normalized F1 values for the low vowels were entered into a repeated-measures analysis of variance with Block and Vowel repeated across subjects. Vowel returned as a significant main effect [$F(1,7)=15$, $p>0.001$], which indicated the fact that /æ/ has a higher F1 than /ɑ/ for this group of participants. There were no other significant effects or interactions. Crucially, participants did not significantly lower their F1 in the course of repeating the words. While this experiment is not directly comparable to the current study because of differences in production between read speech and non-read speech, these results suggest that the process of lowering F1 for the low vowels in the current experiment is due

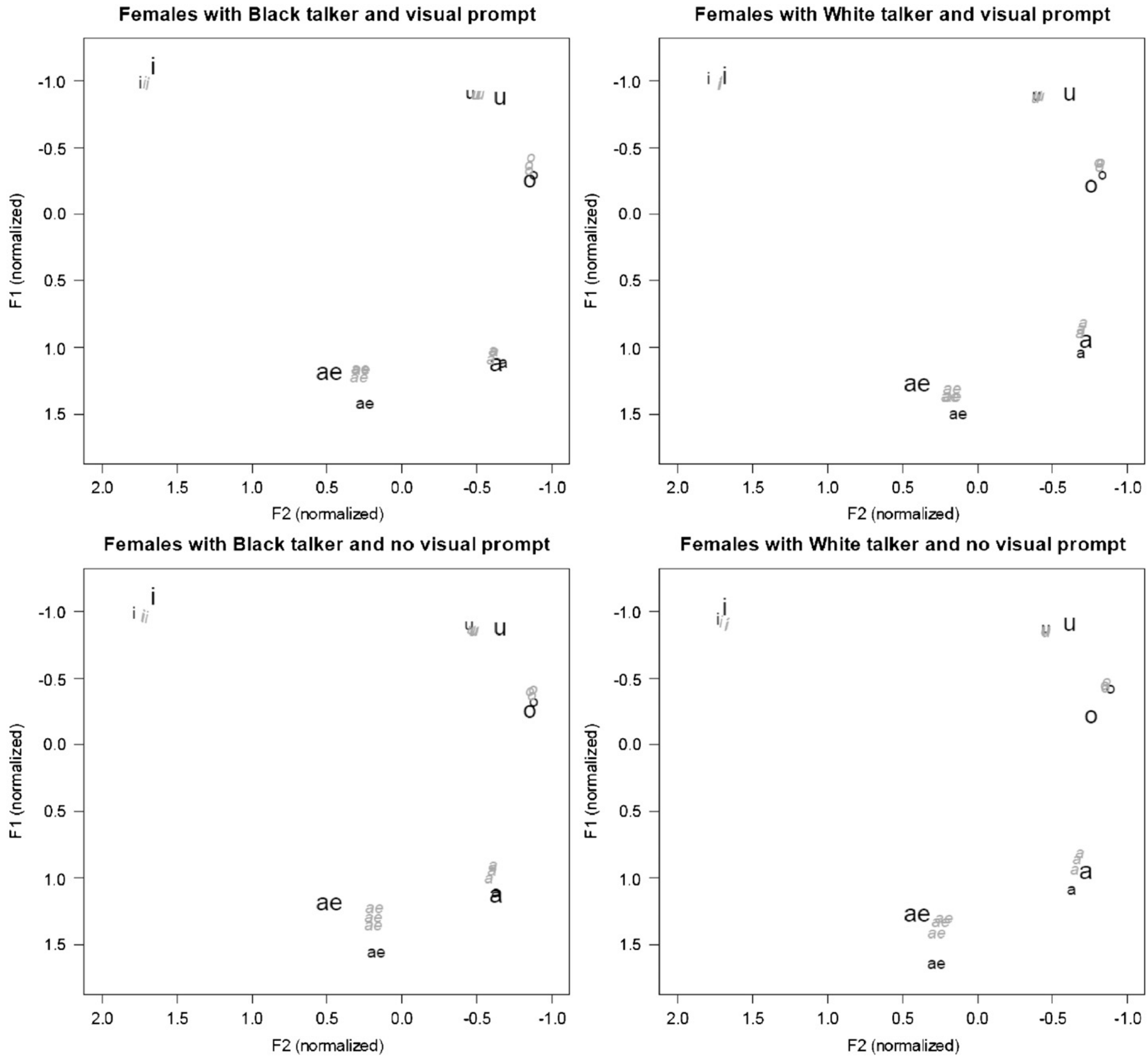

Fig. 6. Formant plot displaying the direction of spontaneous phonetic imitation for the female participants. Normalized formant values are plotted. The mean of the model talkers' vowels are in the slightly larger font and in black. Participants' pre-task vowel means are plotted in a smaller black font and their productions from the three shadowing blocks are in the small italicized grey print.

4.2. Attractiveness: Black and White Visual Prompt Conditions

How attractive participants found the model talker was predicted to globally affect their degree of imitation across the entire shadowing task. Therefore, average distance in distance values were calculated for each participant. Attractiveness rating data were used in an analysis, which took the subset of averaged difference in distance values for each participant for the Visual Prompt Conditions. These values were fit to a linear model along with each participant's attractiveness rating from the model talker they were exposed to, the Gender of the participants (male or female), and the model talker (Black or White talker) they were presented with. So as to not assume attractiveness ratings would be uniform across participant genders and for the two model talkers, participant gender and model talker identity were included as predictors. This model is summarized in Table 3. While the model returned significant interactions with the attractiveness ratings, the overall model presented only a trend toward significance [$F(4, 49)=2.1, p=0.09$]. These data are presented in Fig. 8, illustrating the interactions.

To further explore these relationships, correlation analyses were conducted with participants' averaged difference in distance values and attractiveness ratings by gender and model talker. These analyses revealed no significant relationships for the Black model talker. For the White model talker, however, female participants' difference in distance scores and attractiveness ratings were highly negatively correlated [$t(13)=-2.83$, $p<0.05$, Pearson's

(*footnote continued*)
to auditory exposure to a talker who produced vowels with a lower F1 than most of the participants.

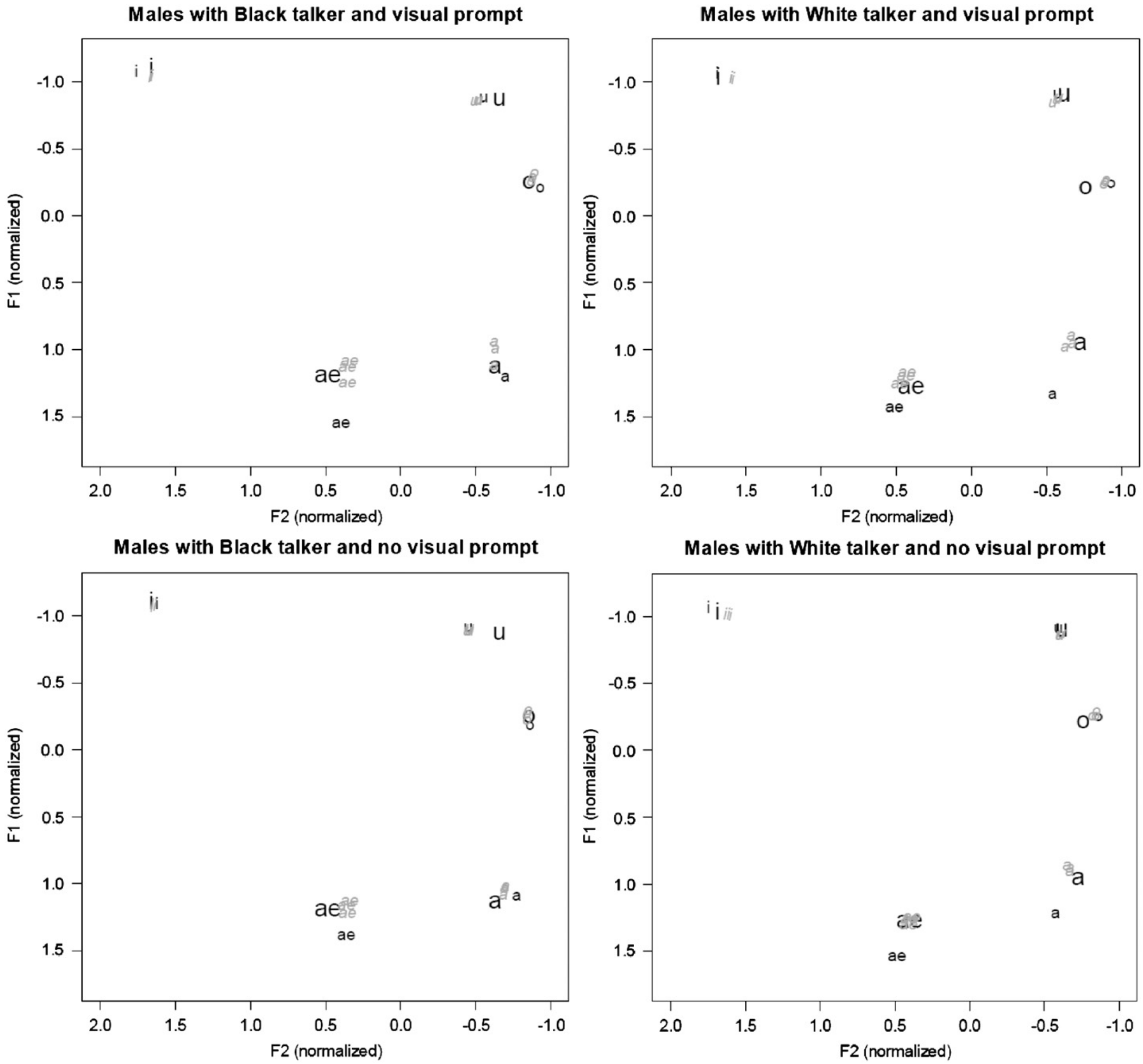

Fig. 7. Formant plot displaying the direction of spontaneous phonetic imitation for the male participants. Normalized formant values are plotted. The mean of the model talkers' vowels are in the slightly larger font and in black. Participants' pre-task vowel means are plotted in a smaller black font and their productions from the three shadowing blocks are in the small italicized grey print.

$R = -0.62$] and men's were strongly positively correlated [$t(10) = 3.27$, $p < 0.01$, Pearson's $R = 0.72$]. For female participants, the more attractive they rated the White model talker, the more likely they were to imitate his vowels. The reverse pattern was found for the male participants: the more attractive male participants rated the White talker, the less likely they were to accommodate toward his vowels. No significant relationship was found for the Black model talker.

5. Discussion

It has been argued previously that phonetic convergence is a socially mediated process (Giles, 1973; Giles & Coupland, 1991; Shepard et al., 2001; Pardo, 2006). Under this view, imitation and convergent speech behaviour are social acts that talkers use to modulate social distances in communication. In the experiment reported here there were several results that can be considered further evidence that social factors mediate the extent of imitation.

Table 3
Results of a linear model where attractiveness ratings, Talker race (Black or White model talker), and participant gender were fit to average difference in distance values.

	β	t-value
Intercept	−0.049	−0.532
Attractiveness rating	−0.02	−1.04
Attractiveness rating: Talker race	0.06	2.03*
Attractiveness rating: Gender	0.041	2.34*
Attractiveness rating: Talker race: gender	−0.09	−2.65*

* $p < 0.05$.

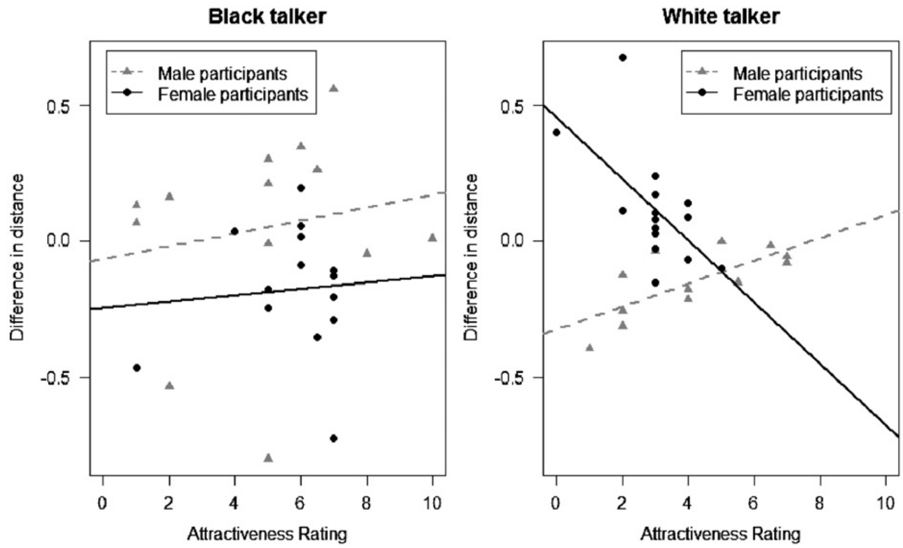

Fig. 8. Attractiveness ratings and averaged difference in distance values. The difference in distance measure on the *y*-axis indicates the amount of phonetic imitation. A value of zero shows no change in vowel production as a result of auditory exposure to the model talkers. A negative value demonstrates phonetic imitation and a positive value demonstrates vocalic divergence. Participants' attractiveness rating for the model talker is on the *x*-axis. A higher rating is more attractive than a lower rating. Data points for female participants are shown as black circle and the regression line for female's is the solid black line. Male data points are grey triangles and the regression line for the male data is the dashed grey line.

Moreover, the tendency to imitate low vowels more than /i o u/ is evidence that phonetic convergence may be conditioned by phonetic factors as well. As is discussed below it is not clear, however, that separating the social from the phonetic is a feasible task.

More imitation occurred in the Visual Prompt Condition, lending support to the hypothesis that convergence is a socially driven phenomenon. Due to the complexity of the experimental design, the Visual Prompt and No Visual Prompt Conditions were then separated to avoid interpreting unwieldy and overly complex interactions between dependent variables. Prior to this division in the dataset, however, the first analysis revealed a relationship between the presence of a picture of the talker, the model talker's voice, participant gender, experimental block, and vowel. What is important about this interaction is that it did not simply involve Picture and Voice—participants' pattern of response to the stimuli was much more complex than simply being a reaction to the identity of the model talker.

Within the No Visual Prompt data where less imitation was found overall, there were no effects of Voice, and, in general, the results in those conditions are much more straightforward compared to the Visual Prompt data. The Visual Prompt Condition, in presenting an image of what the model talker looks like, is arguably more socially engaging. Dijksterhuis and Bargh (2001) suggest that all types of behavioural imitation are automatic, but that social knowledge can inhibit or facilitate the degree of imitation. In the case of spontaneous phonetic imitation, social information about the model talkers garnered from the visual signal resulted in more imitation, as compared to the No Visual Prompt Condition.

These results suggest that the variable nature of imitation found in the more socially meaningful context is likely due to various social influences. Participant gender was the only social factor that affected imitation in the No Visual Prompt Condition. Overall, the No Visual Prompt Conditions found that /æ/ was imitated more than other vowels, but that female participants were primarily responsible for this effect. The Visual Prompt data presented a much more complex picture with interactions involving gender, vowel, and model talker voice. Males in the White talker Visual Prompt Condition accommodated more to the model's /ɑ/, whereas in all other conditions, /æ/ was the vowel that exhibited the most imitation. Note that it was generally not the case that /i o u/ were not imitated, but that /æ/ and /ɑ/ were imitated more consistently and to a greater extent. Why then were these vowels imitated more than the others? The low vowels /æ/ and /ɑ/ are vowels that exhibit considerable regional variation in North American English (Clopper, Pisoni, & de Jong, 2005). It could be the case that these vowels were imitated to a greater extent because participants came to the task with low vowels that differed more from those of the model talkers, and therefore participants had acoustic-phonetic space in which to accommodate. Unfortunately, participants' dialect histories were not explicitly collected in this study; the vast majority of participants were born and raised in California, but this was not true for all.[2] In post-hoc explorations of the effect of dialect on /ɑ/ imitation for males in the White talker Visual Prompt Condition, a curious discovery was made. The three participants who exhibited the greatest amount of imitation (out of all of the participants in the entire study) were three male participants born and raised in the Upper Midwest who had all been randomly assigned to the White talker Visual Prompt Condition. Formant plots for these three participants are presented in Fig. 9. Like in the previous formant plots, in these figures the average model talker vowels are plotted in the larger black font, the participants' averaged baseline productions are in the smaller black font, and their averaged shadowed productions across the three shadowing blocks are in grey italics. This figure makes it clear that not only was there significant imitation in /ɑ/, but that this is the vowel that started out as most distinct for these participants compared to the model. This finding echoes recent cross-dialectal accommodation work by Babel (2010), which examined the patterns of imitation by New Zealanders in an auditory shadowing task with an Australian model talker. Participants were found to accommodate most towards the DRESS vowel—a vowel that occupies rather different parts of the vowel space for New Zealand and Australia talkers. Interestingly, while this is one of the largest differences in New Zealand and Australian Englishes, New Zealanders are less aware of this dialect difference (Hay, Nolan, & Drager, 2006). It is well established that position of

[2] Data was collected on where participants had previously lived, but these data proved impossible to categorize succinctly in such a way that they could be incorporated as a statistical variable into the study.

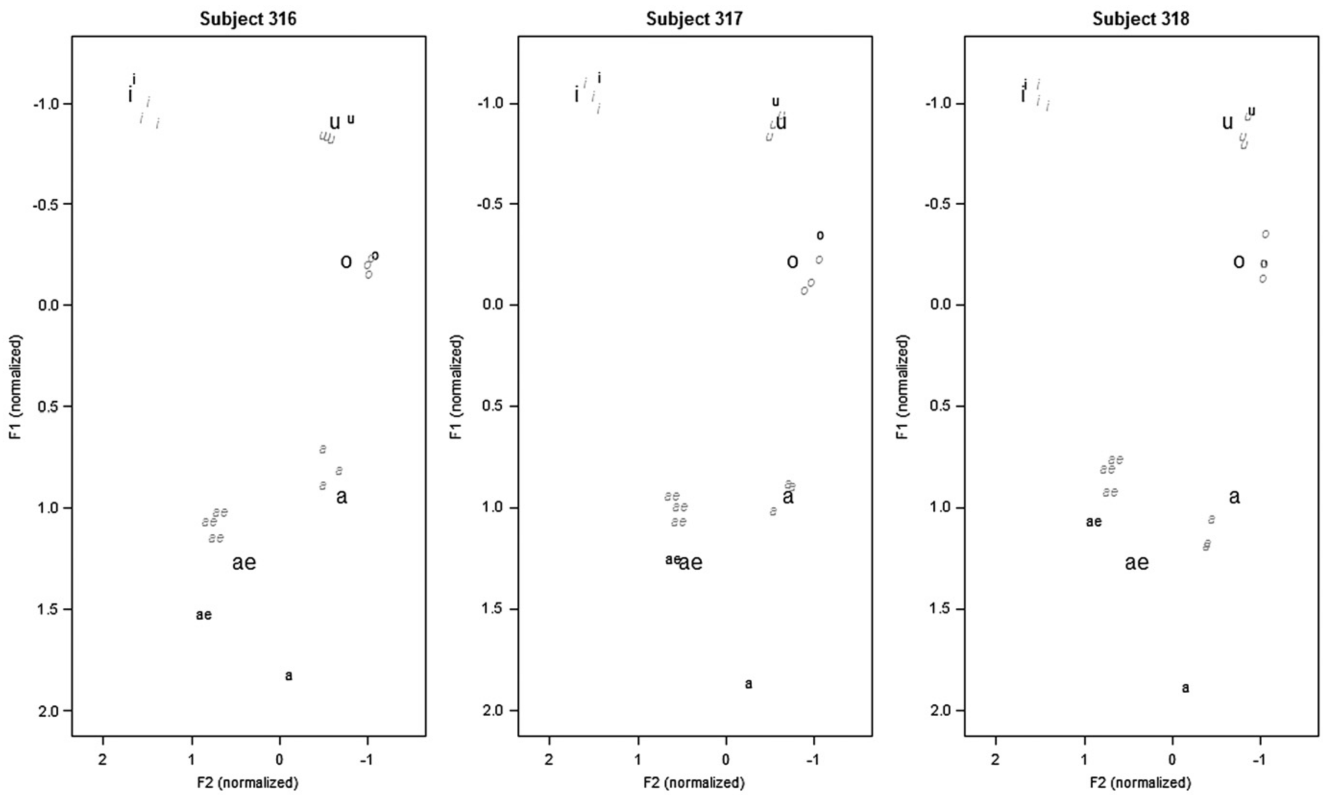

Fig. 9. Formant plots from the three participants who exhibited the most imitation. Formant plot displaying the direction of spontaneous phonetic imitation for the male participants. Normalized formant values are plotted. The mean of the model talkers' vowels are in the slightly larger font and in black. Participants' pre-task vowel means are plotted in a smaller black font and their averaged productions from all three shadowing blocks are in the small italicized grey print.

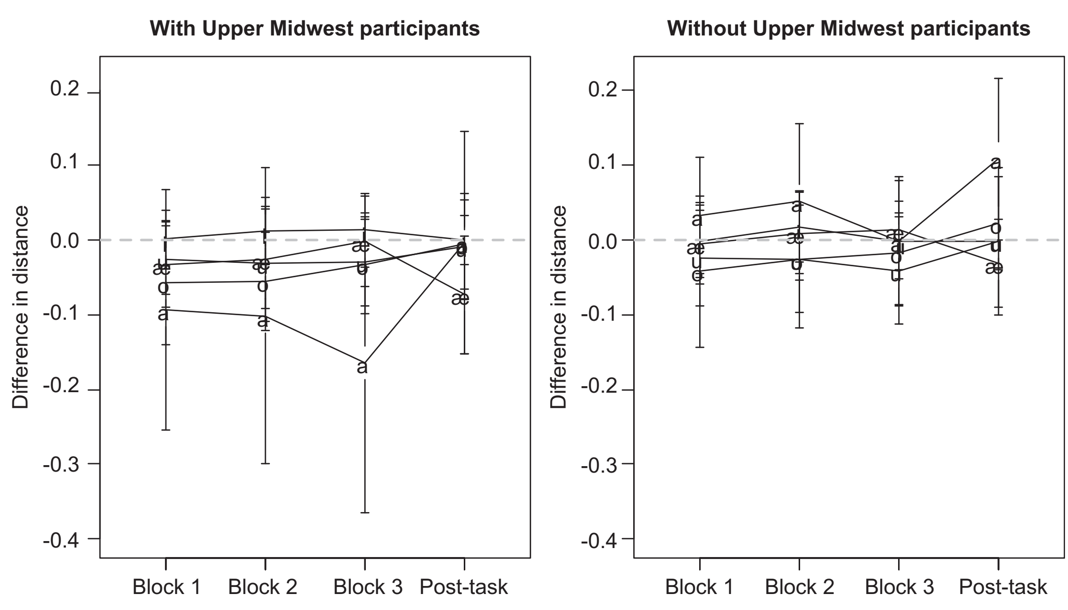

Fig. 10. The original data from the White talker Visual Prompt Condition for male participants with the Upper Midwest participants (left panel) and without these participants (right panel). The difference in distance measure on the *y*-axis indicates the amount of phonetic imitation. A value of zero 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. Task Blocks 1, 2, and 3 are Shadowing Blocks; the Post-task Block is labelled as such. The error bars represent 95% confidence intervals.

low vowels is a defining characteristic of these participants' Upper Midwest dialect, but whether these differences are salient to speakers of the variety is unknown. Regardless, these results indicate that dialect background matters for degree of imitation (see also discussion in Vallabha and Tuller (2004) on the role of dialect in explicit steady-state vowel imitation); greater phonetic distance, which is a function of dialect background, seems to allow for more imitation. Fig. 10 compares the data of male participants in the

White talker Visual Prompt Condition with (left panel) and without (right panel) the three Upper Midwest participants. In comparing these panels, it is clear that the effect of /ɑ/ being imitated more in this condition was driven by these three male participants.

A goal of this experiment was to offer a test-case for exemplar-based models of speech production and perception. In such models, exposure to a production activates similar stored exemplars, leading in turn to a subsequent production by the listener-turned-talker that is a more like that of the model talker. Such models are able to account for convergent results, if an adaptable model that allows for attention weighting is implemented (e.g. Goldinger & Azuma, 2003; Johnson, 1997), but they do not provide a straightforward explanation for the cases of divergence. Moreover, the lack of equal imitation in all conditions indicates that a simple automatic exemplar-based model is untenable and that the amount of activation or the amount of inhibition in the process of selection for production must interface with basic social desires such as liking. Additionally, exemplar-models predict that more exposure would increase activation, making the imitation results in this experiment cumulative across shadowing blocks. The results, however, provide evidence that increased exposure to the model talker does not cause a significant increase in imitation. This finding supports theories in the syntactic priming literature that accommodation phenomena involve surprisal and are not cumulative (Jaeger and Snider, 2007).

Finally, Dijksterhuis and Bargh (2001) argue that the strength of the perceptual-behaviour link is mediated by disincentives and obstacles that conflict with goals, attention, and liking. Liking was measured in this experiment by asking participants to rate the attractiveness of the model talker to which they were exposed. Female and male participants behaved rather differently with respect to accommodation and liking and this was affected by the model talker as well. In the White talker Visual Prompt Conditions, the more attractive females rated the model talker, the more likely they were to accommodate his low vowels. These data align with predictions from CAT, which suggest that when a talker is viewed positively, there is a desire to decrease social distance and this will be, partially, accomplished by imitating. Male participants, on the other hand, did not perform as predicted by CAT. The more attractive male participants rated the White model talker, the less likely they were to accommodate to his vowels. One possibility for this divergent behaviour is to interpret the male divergence as a response to a threat. Male participants who rated the male model talker as attractive were, perhaps, socially threatened and distanced themselves in response to the threat. Another possibility is that asking male participants to rate the attractiveness of a male model talker does not provide the same measurement as asking a female participant to do the same.[3]

The interaction between amount of imitation and liking is important for sociolinguistic theories of style- and dialect-shifting. While many such theories remain agnostic with respect to the role of consciousness in the style-shifting process (Bell, 1984, 2001), some scholars have claimed that such variation is under a talker's control. For example, Labov (2001: 85) has claimed style is a "controlled device" in language use, while Eckert (2001: 124) has described language has a "fairly low level process" with its social meaning being subject to "conscious manipulation". These results suggest that at least some of the motivation for style- and dialect-shifting stems from low-level imitation that is mediated by implicit social factors such as liking (Babel, 2010).

[3] The sexual orientation of participants was not collected and this interpretation can only be safely posited with the assumption that the majority of male participants in this study were straight.

6. Conclusions

The results reported here suggest that phonetic imitation is selective from both a phonetic and social perspective. Not all vowels were imitated to the same degree—it seems that there is more imitation when there is the phonetic space to do so. To this end, it seems that participant dialect may play a crucial role in accommodation, which may provide the impetus for dialect acquisition and change (Trudgill, 2008). Additionally, the fact that more imitation occurred in the condition that involved a visual image of the model talker suggests that social context facilitates the process.

These data support an automatic theory of accommodation that assumes implicit social factors will play a role on an unintentional and subconscious level (e.g. Pickering & Garrod, 2004). However, since imitation is not a necessary consequence of auditory exposure, it is unlikely that production and perception work out of the same representational stores. It should be noted, though, that the fact that auditory exposure can shift production targets in the first place implies a strong connection between the two processes. A major issue not addressed in this research is whether the failure to accommodate stems from the inability to resolve perceptual details, biases in production, or inattention to the stimulus; determining the roles of each of these possibilities is left to future investigation.

With respect to representation, these data provide evidence against the claim made by Mitterer and Ernestus (2008) that imitation is mediated by phonological representations. Instead, the results indicate that gradient acoustic information encountered in auditory processing carries over into the fine details of speech production. Vowel accommodation in spontaneous phonetic imitation 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, the results reported here demonstrate that part of the variability in speech production can be accounted for by auditory exposure during or prior to production.

Acknowledgements

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. This manuscript has been greatly improved by the feedback of three anonymous reviewers.

References

Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America*, *116*, 99–107.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics for R*. Cambridge: Cambridge University Press.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database (CDROM). Linguistic data consortium*. University of Pennsylvania.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*, 437–456.

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*, 145–204.

Bell, A. (2001). Back in style: Reworking audience design. In P. Eckert, & J. R. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 139–169). Cambridge: Cambridge University Press.

Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*, 355–387.

Boersma, P., & Weenink, D. (2005). *Praat: Doing phonetics by computer (version 4.3.29) [computer program]*. Amsterdam: Institute of Phonetic Sciences.

Bourhis, R. Y., & Giles, H. (1977). The language of intergroup distinctiveness. In H. Giles (Ed.), *Language, ethnicity, and intergroup relations* (pp. 119–136). London: Academic Press.

Branigan, H. P., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition*, *104*, 163–197.

Bryne, D. (1971). *The attraction paradigm*. New York: Academic Press.

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

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic cues. *Cognition*, *108*, 804–809.

Clopper, C., Pisoni, D., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America*, *118*, 1161–1676.

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

Dijksterhuis, A., & Bargh, J. A. (2001). The perception–behavior expressway: Automatic effects of social perception on social behavior. In M. Zanna (Ed.), *Advances in experimental social psychology* (pp. 1–40). San Diego: Academic Press.

Eckert, P. (2001). Style and social meaning. In P. Eckert, & J. R. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 119–126). Cambridge: Cambridge University Press.

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

Fowler, C. (1996). Listeners hear sounds, not tongues. *Journal of the Acoustical Society of America*, *99*, 1730–1741.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination, and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, *53*, 181–215.

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

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

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

Goldinger, S. D. (1997). Perception and production in an episodic lexicon In K. Johnson, & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 33–66). San Diego: Academic Press.

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

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

Goldinger, S. D., & Azuma, T. (2004). Episodic memory in printed word naming. *Psychological Bulletin Review*, *11*, 716–722.

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

Gregory, S. W., Dagan, K., & Webster, S. (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, S. W., Green, B. E., Carrothers, R. M., & Dagan, K. A. (2001). Verifying the primacy of voice fundamental frequency in social status accommodation. *Language Communication*, *21*, 37–60.

Gregory, S. W., & Hoyt, B. R. (1982). Conversation partner mutual adaptation as demonstrated by Fourier series analysis. *Journal of Psychological Research*, *11*, 35–46.

Gregory, S. W., & Webster, S. (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.

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

Harrington, J. (2006). An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics*, *34*, 439–457.

Harrington, J. (2007). Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. In J. Cole, & J. Hualde (Eds.), *Laboratory phonology 9: Phonetics and phonology* (pp. 125–144). Berlin: Mouton de Gruyter.

Harrington, J., Palethrope, S., & Watson, C. I. (2000a). Does the Queen speak the Queen's English? *Nature*, *408*, 927–928.

Harrington, J., Palethrope, S., & Watson, C. I. (2000b). Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association*, *30*, 63–78.

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, *23*, 351–379.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*, 458–484.

Howell, P., Barry, W., & Vinson, D. (2006). Strength of British English accents in altered listening conditions. *Attention, Perception, & Psychophysics*, *68*, 139–153.

Jaeger, T.F. & Snider, N. (2007). Implicit learning and syntactic persistence: Surprisal and cumulativity. In: Wolter, V., & Thorson, J. (Eds.), University of Rochester working papers in the language sciences, Vol. 3(1), pp. 26–44.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson, & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego: Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *24*, 485–499.

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

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal?. *Cognitive Psychology*, *51*, 141–178.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin Review*, *13*, 262–268.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*, 1–15.

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

Kuhl, P., & Metzloff, A. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, *100*, 2425–2438.

Labov, W. (2001). The anatomy of style-shifting. In P. Eckert, & J. R. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 85–108). Cambridge: Cambridge University Press.

Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, *29*, 117–123.

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

Lobanov, B. (1971). Classification of Russian vowels spoken by different listeners. *Journal of the Acoustical Society of America*, *49*, 606–608.

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

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from a shadowing task. *Cognition*, *109*, 68–173.

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

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

Natale, M. (1975a). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, *32*, 790–804.

Natale, M. (1975b). Social desirability as related to convergence of temporal speech patterns. *Perceptual 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.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*, 132–142.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–238.

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

Pardo, J. S. (2009). Expressing oneself in conversational interaction. In E. Morsella (Ed.), *Expressing oneself/expressing one's self: Communication, cognition, language, and identify* (pp. 183–196). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, *134*, 427–459.

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

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

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

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

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

Shepard, C. A., Giles, H., & LePoire, B. A. (2001). Communication accommodation theory. In W. P. Robinson, & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 33–56). New York: John Wiley & Sons Ltd.

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

Tilsen, S. (2009). Subphonemic and cross-phonemic priming in vowel shadowing: Evidence for the involvement of exemplars in production. *Journal of Phonetics*, *37*, 276–296.

Trudgill, P. (1986). *Dialects in contact*. New York: Blackwell Publishing.

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

Vallabha, Gautum K., & Tuller, Betty. (2004). Perceptuomotor bias in the imitation of steady-state vowels. *Journal of the Acoustical Society of America*, *116*, 1184–1197.