Phonetic Convergence Across Regional Dialect: the Roles of Dialect-Specific Variation and Talker Dialect Distance

by

Rebecca Piper

Readers: Dr. Susannah Levi and Dr. Lisa Davidson

Written in partial fulfillment of the requirements for departmental honors in Language and Mind

New York University

April 1, 2015

Abstract

Phonetic convergence is the phenomenon in which people unintentionally and temporarily change phonetic details of their speech to sound more similar to another talker. Previous research has revealed multiple lexical/task and social factors that affect the presence and degree of phonetic convergence. Of interest to the current study is the influence of regional dialect distance, which is the similarity of the dialect of the speaker who converges (the shadower) and of the other talker (the target talker). It is unclear whether more or less similarity between the dialects of the shadower and target talker results in stronger phonetic convergence. In the current study, shadowers with a General American dialect repeated words recorded by two target talkers, one with a General American dialect and the other with a New York City dialect. Results of a perceptual discrimination task revealed that speakers were more likely to modify their speech when listening to the General American talker, supporting previous research that speakers are more likely to converge when there is greater similarity between their own dialect and the dialect of the target talker.

Introduction

As listeners, we are exposed to an enormous variety of acoustic information in the speech signal. Not only do different talkers produce sounds differently as a result of physiological characteristics – larger vocal tracts will result in speech that resonates at lower frequencies, for example – but individual acoustic variation also exists as a result of group-specific qualities such as regional dialect differences. Previous research has shown that people in fact modify their speech to match the phonetic variations in other talkers' speech. This modification has been found to affect several acoustic properties such as fundamental frequency (f0) and f0 trajectory (Goldinger, 1998; Pardo, Jay, & Krauss, 2010; Phillips & Clopper, 2012), intensity (Natale, 1975), word duration (Goldinger, 1998; Pardo et al., 2010), vowel quality (Babel, 2009, 2010, 2012), and voice onset time (VOT) (Shockley, Sabadini, & Fowler, 2004). These *phonetic convergence* studies tell us a great deal about our ability to perceive subtle details of spoken language and how information encoded during perception gets translated during production.

Interestingly, it has also been found that changing your speech to sound more like the speech of another talker has significant effects on the quality of interaction between interlocutors. Adank, Hagoort, and Bekkering (2010) investigated the effects of overt, or purposeful, imitation on sentence comprehension. Participants were initially tested on their ability to repeat sentences spoken in noise with an unfamiliar accent. Following pre-test, they were divided into groups and received varying types of training on a different set of sentences spoken with the same unfamiliar accent. Training conditions included a contrast between non-imitative repetition and overt imitation where one group was instructed to repeat the sentence spoken by the target talker but to do so in their own accent, while another group was told to repeat the sentence while purposefully imitating the talker's accent. They were then tested again

on repetition of the original test sentences in noise. Post-test results revealed that the participants who purposefully imitated the target talker's accent were able to repeat 50% of the key words in each sentence in higher levels of background noise than the participants who were not told not to imitate. Adank et al. argued that imitation of another talker facilitated speech comprehension.

In another study contrasting overt imitation and non-imitative repetition, Adank, Stewart, Connell, and Wood (2013) investigated how convergence affected shadowers' judgments of target talker attractiveness. In their study, British English - speaking participants listened to sentences recorded by two different target talkers both speaking Glaswegian English, a non-standard dialect spoken in the region surrounding Glasgow, Scotland. For the imitation task, they used the same instructions as Adank et al. (2010) to elicit non-imitative repetition and overt imitation: each participant was instructed to repeat the utterances of one target talker *without* imitating the dialect and to purposefully imitate the other target talker. Following the repetition of each target talker, participants rated them on various traits pertaining to voice quality, personality, and social, educational, and economic status. Results revealed that participants rated the target talker that they had overtly imitated higher in social attractiveness than the target talker they had simply repeated, indicating that imitation also affects shadower attitudes toward the target talker.

While it is important to note that the results of the Adank et al. (2010, 2013) studies only revealed effects of overt imitation and not the spontaneous convergence studied in this paper, and therefore might not carry over to more subtle forms of convergence, they still demonstrate the importance of imitation of a talker's speech in interlocutor interactions. By studying spontaneous phonetic convergence, we can better understand both the conscious and unconscious

mechanisms behind shifts in speaking style and apply them to the changes that take place during every day communication and long-term dialect change.

Research has revealed the influence of a wide range of factors on how we change our speech in response to the speech of another talker. While it is clear that certain social and lexical/task factors such as social appeal, gender, talker role, lexical frequency, and exposure affect the degree to which talkers converge, it remains unclear whether greater similarity or greater difference between two talkers results in greater phonetic convergence. In the current study, speakers were provided an opportunity to converge to both a similar dialect speaker and a different dialect speaker, addressing the question of whether dialect similarity affects phonetic convergence.

Phonetic Convergence

The phenomenon where people unintentionally and temporarily change phonetic details of their own speech to be more similar to the characteristics of another person's speech is called *phonetic convergence*. Studies have found phonetic convergence in both socially-rich contexts and also in situations that are socially-minimal. First, socially-rich approaches to studying phonetic convergence have used interactive, conversation-based tasks to elicit phonetic convergence (Kim, Horton, & Bradlow, 2011; Pardo, 2006). These studies use paradigms designed to create more spontaneous conversational speech samples as well as to provide participants with a common goal during interaction. In one such study, Pardo (2006) employed a map task, in which two participants were each given a map which the other could not see. One map had more detailed information, and this participant (the giver) was instructed to guide the other participant (the receiver) to some destination on the map by following a specific set of

landmarks. To assess convergence, all partners' productions were compared across the duration of the task. For example, the giver's productions of a particular landmark (e.g., "bank") during the task were compared to two items: the receiver's repetition of the same landmark name during the task as well as the giver's pre-task production of the landmark. The giver's post-task production was also compared to measure the perseverance of convergence beyond completion of the task. Pardo found that both the participants' within-task repetitions and their post-task utterances were more similar to their partners' within-task utterances than were their pre-task repetitions, showing that they had converged to their conversation partner over the course of the experiment and that the effect lasted beyond the task. More recently, Kim, Horton, & Bradlow (2011) used a naturalistic *diapix conversation elicitation* paradigm to create even more naturalistic utterances. In their study, two participants were each given a drawing, which the other could not see, of the same scene with 10 differences between the two drawings. They were instructed to work with each other to find the differences. To assess convergence, Kim et al. compared participants' utterances at the beginning and the end of the task to their partner's utterances at the end of the task and found that participants' end-of-task utterances were more similar to their partner's end-of-task utterances than were their beginning utterances, showing convergence. What is more, Kim et al. found that certain participant groups converged more than others. Pairs that spoke with the same dialect converged more than pairs that did not speak the same dialect, suggesting that dialect similarity affected the degree to which the partners converged (see below for more detail on this study). Finally, another study that also found convergence in a conversation-based task was Natale (1975), in which interviewer – interviewee pairs seated in separate audio booths (so as to remain invisible each other) were recorded in a task in which the interviewer, in five-second utterances, elicited discourse from the interviewee

about a general world topic. The intensity of interviewer's voice, heard by the interviewee through speakers, was controlled to be divided into three different intensity conditions. Results revealed that rises in the intensity of the interviewer were correlated to the rises in intensity of the interviewees, suggesting that the interviewees had converged to this particular property of the interviewer's speech during the conversation.

Interestingly, phonetic convergence has proven to be a robust effect, having also been found in less natural, non-interactive settings where there appears to be little social motivation for convergence. In the first study of this type, Goldinger (1998) asked participants to first read a set of words which were used as the baseline productions. These participants, called *shadowers*, then completed an auditory naming task, or shadowing task, in which they were asked to identify single words they heard over headphones by saying them into a microphone. Phonetic convergence was assessed by testing whether the shadowers altered their speech to sound more like the talkers they heard, called *target talkers*. To do this, the shadowed productions were presented to a new set of listeners along with the baseline productions and the target talker's productions in an AXB perception task, in which listeners decided which of the baseline or shadowed utterances sounded more like, or like a better imitation of, the target talker utterance (see below for more details on the AXB task). Importantly, the results of Goldinger's perception task revealed that phonetic convergence occurred in the socially-minimal task setting, indicating convergence is to some degree an automatic process not solely driven by social factors.

Several other studies have also used Goldinger's socially-minimal tasks to study phonetic convergence in a lab setting. For example, Shockley, Sabadini, and Fowler (2004) first replicated Goldinger's 1998 experiment, then artificially elongated the VOTs of the word-initial stops of the target talker utterances and presented them in an auditory naming task to a new group of

shadowers. Using acoustic analysis, they found that the shadowed VOTs were significantly longer than the baseline VOTs, showing that VOT is an acoustic property affected by phonetic convergence. In another socially-minimal task, Phillips and Clopper (2012) assessed whether dialect familiarity could affect phonetic convergence. Shadowers from three different dialect groups – Northern Cities, Midland, and Mobile (having lived in more than one dialect region prior to the study) completed auditory naming tasks with six different target talkers, three each from the Northern Cities and Midland dialects. A subset of the baseline and shadowed utterances representing vowels for each of the six target talkers was then presented to listeners who determined convergence by comparing them to the target talker utterances in an AXB listening task similar to that of Goldinger (1998). However, to assess the role of dialect familiarity, they then introduced a novel element to their perceptual listening task, in which listeners compared the baseline and shadowed productions of a single shadower to one of three conditions: the original target talker that the shadower heard (same talker condition), a different target talker with the same dialect (different talker condition), or a target talker with a different dialect (different dialect condition). Phillips and Clopper found that while listeners perceived phonetic convergence above chance only in the same talker condition, the results of the different talker condition were significantly higher than the different dialect condition, showing that shadowers converge on dialect characteristics as well as talker-specific variations.

Finally, phonetic convergence is so robust that it has also been found to occur across modalities, without the presence of audible target utterances. For example, Miller, Sanchez, & Rosenblum (2010) found that phonetic convergence can result from visual speech. In their study, shadowers lip-read silent clips of target talkers and said the words out loud. Even without audible

stimuli, Miller et al. found that the shadowed utterances were perceived to be more similar to the target utterances than were the baseline utterances.

Factors affecting phonetic convergence

Studies of phonetic convergence have revealed several factors that affect the degree of convergence, which are summarized in Table 1.

Category	Factor(s) affecting convergence
Social	racial/ethnic bias attractiveness & uniqueness gender relationship quality conversation role dialect distance
Experimental task/stimuli	word frequency amount of exposure dialect distance

Table 1. Summary of factors found to affect phonetic convergence. Dialect distance, defined later in the section, can be considered both social and task-related.

To begin, some of the factors that affect phonetic convergence are socially motivated. For example, Babel (2010) implemented an auditory naming task with words containing the vowels /i, æ, u, o, ɑ/, in which the target talker was from Australia and the shadowers were from New Zealand. Following the task, shadowers completed an Implicit Association Task, which used differences in reaction time during categorization tasks to indicate either a pro-Australia or pro-New Zealand bias. Acoustic analysis of the target talkers' and shadowers' vowel qualities showed that shadowers who demonstrated a pro-Australia bias were more likely to converge to the Australian target talker, indicating that social biases played a role in convergence. In addition, Babel (2009, 2012) used similar methodology to show that shadowers who

demonstrated a pro-black racial bias in an Implicit Association Task converged more to a black target talker than did shadowers with less of a pro-black bias. From these studies, Babel argued that social preferences help determine the type and degree of communicative accommodation. In addition to exploring cultural biases, Babel (2009) also found that degree of phonetic convergence was greater toward target talkers previously rated to be most attractive and least typical, showing that judgments of attractiveness and uniqueness also influence convergence.

Another social factor found to affect convergence is gender. Namy, Nygaard, & Sauerteig (2002) used a standard shadowing task and AXB listening test and found that overall, female shadowers converged more than male shadowers, indicating that shadower gender affects phonetic convergence. Similar results were also found in Babel, McGuire, Walters, & Nicholls (2014), in which female shadowers showed more phonetic convergence than male shadowers overall. The results of the conversational task in Pardo (2006), however, showed that males converged more than females. Pardo (2006) also found that for male participant pairs the participant who received instructions from his partner to complete the map (the receiver) converged more than the participant who gave the instructions (the giver). More recently, Pardo (2010) replicated the 2006 map task study but with one change: in every pair, one conversation partner, counterbalanced between giver and receiver, was explicitly instructed to imitate the other partner's speech. She found that even when the receivers were explicitly told to imitate, the givers were still judged to have converged more to receivers than the other way around, and receivers were not judged to have converged to givers at all. These studies revealed that conversation role also affects phonetic convergence. Finally, Pardo (2012) found that convergence between randomly-assigned (male) college roommate pairs over months of living

together was correlated with the roommates' self-reported closeness, suggesting that the quality of the relationship between interlocutors could have an effect on convergence.

In addition to socially-motivated factors, features of the lexicon also affect the presence and degree of phonetic convergence. First, studies have shown that lexical frequency plays a role. For example, Goldinger (1998) found that low frequency words (less than 75 per million (Kucera & Francis, 1967)) elicit greater convergence than high frequency words. Convergence has also been found for low frequency words by Goldinger and Azuma (2004), Babel (2010), and Nielsen (2011). Second, Goldinger (1998) also found that shadowers are more likely to converge to words after hearing the word spoken by the target talker multiple times, suggesting that phonetic convergence is influenced by amount of exposure. Most subsequent studies using the shadowing paradigm have thus used only low frequency words and multiple repetitions (Babel, 2012; Miller et al., 2010; Namy et al., 2002; Shockley et al., 2004). However, see Pardo, Jordan, Mallari, Scanlon, and Lewandowski (2013) for an argument for the inclusion of high frequency words in studies of phonetic convergence.

An additional factor that has received attention recently, and the one that is the most relevant for the current study, is the *dialect distance,* or dialectal similarity of the language, of the target talker and the shadower. Interestingly, the four studies detailed below show conflicting results: in one, greater similarity resulted in phonetic convergence, in two greater differences resulted in phonetic convergence, while the fourth showed inconclusive results. As discussed above, Kim, Horton, and Bradlow (2011) tested the amount of phonetic convergence using a naturalistic diapix conversation elicitation paradigm in which each member of a conversation pair was given a slightly different version of a picture and instructed to work with each other to find the differences between their pictures. To investigate the effects of language similarity, they

tested perceived phonetic convergence of early- and late-task utterances in three types of conversation pairs: same L1/same dialect, same L1/different dialect, and different L1. To measure phonetic convergence, they presented the early utterances, partner utterances, and late utterances to listeners in a modified AXB task (XAB) in which listeners heard the partner utterances first followed by the early and late utterances, which were counterbalanced. Results revealed that the same L1/same dialect group showed more phonetic convergence than the other two, suggesting that more similarity between talkers allows for greater convergence. Furthermore, there was no difference between the degree of convergence between the same L1/different dialect and different L1 groups, suggesting that, unlike sharing a dialect, sharing an L1 does not promote convergence compared to not sharing an L1.

In contrast to Kim et al. (2011), Babel (2012) found that larger differences between talkers resulted in more convergence. She used a standard shadowing paradigm to examine phonetic convergence of vowels, using speakers of California English as the target talkers. She found that the shadowers who showed the greatest change in production of the low vowels /a/ and /æ/ toward that of the target talker were those born and raised in the Upper Midwest, where the Northern Cities dialect is spoken. For speakers of the Northern Cities dialect, the low vowels are known to be significantly shifted from General American English (Labov, Ash, & Boberg, 2006). Based on these findings, Babel argued that shadowers whose dialect is less similar to the target talkers' are more likely to show phonetic convergence because they have more room in their phonetic space to converge.

Walters, Babel & McGuire (2013) presented participants with the same shadower baseline and target talker productions that would be used in Babel et al. (2014) (detailed above), which found significant effects of convergence, and asked them to rate their similarity using a

Visual Analog Scale. They then compared the similarity ratings to the results of the AXB test from Babel et al. (2014). Results revealed that the baseline voices that were rated as more different from the target talker voices on the Visual Analog Scale showed more perceived phonetic convergence in the prior study. Interestingly, this effect was only found for female shadowers, while male shadowers showed no correlation between language similarity and phonetic convergence. Walters et al. used the results of this study to confirm the results of Babel (2012) and argue against the results found by Kim et al. (2011). However, it is important to note that tests of utterance similarity do not necessarily reflect dialect distance, rather, they reflect voice similarity in general. Therefore, the results reported by Walters et al. (2013) may not reflect only dialect distance. To measure dialect distance specifically, it is necessary to control specifically for the dialects of the target talkers.

Finally, Phillips and Clopper (2012), in addition to studying whether shadowers were sensitive to regional dialect (see above for more detail on this study), assessed whether dialect familiarity, measured by region of origin, was a factor in convergence. Their study used one set of target talkers from the Midland dialect region and another from the Northern Cities region. Their shadowers were Midland, North or Mobile (having lived in more than one dialect region). Phillips and Clopper predicted that shadowers would respond to less familiar dialects similarly to the way they have been shown to respond to low frequency words: they would be more likely to converge to target talkers with less familiar dialects, possibly as a result of having fewer representations of that dialect in memory. For example, it was expected that Northern shadowers would converge less for Northern target talkers than would Midland shadowers. However, results of acoustic analysis of convergence in vowel quality, f0, f0 trajectory, and coda duration were

inconclusive with regard to the effects of dialect familiarity. The conflicting results with regard to the effects of regional dialect make it an intriguing topic to explore in the current study.

Theories of phonetic convergence

Two distinct categories of theories for predicting the mechanisms and motivation behind phonetic convergence have been proposed: social and automatic. First, a prominent socially-motivated theory of phonetic convergence is Communication Accommodation Theory (CAT) developed by Giles and colleagues (Giles, 1973; Giles, Mulak, Bradac, & Johnson, 1987). According to CAT, an individual's strategic, social-psychological desire to manage social relationships, gain social approval, and establish social identity predicts when one member of a conversation pair is likely to converge to his partner, diverge from (sound even less like) his partner, or carry out maintenance (change nothing). Under this account, conversation partners converge to minimize social distance and diverge when they want to emphasize their differences.

Phonetic convergence research has provided some degree of support for theories attributing convergence to social factors. For example, Babel (2010) (detailed above), in which New Zealand shadowers converged to the vowels of an Australian target talker, showed that shadowers whose Implicit Association Task revealed a pro-Australia bias converged more to the target talkers than did shadowers with a pro-New Zealand bias. The results of this study and of Babel (2009, 2012) (also detailed above), in which it was found that shadowers with a pro-black bias converged more to black target talkers than did shadowers with less of a pro-black bias, indicate that phonetic convergence is influenced by the social preferences of the shadower.

The second major category of phonetic convergence theory holds that convergence is a much more automatic process. An example of a less-socially motivated, more automatic theory

that accounts for phonetic convergence is exemplar theory (Goldinger, 1998). In this theory, phonetic convergence results from a more direct relationship between perception and production. It argues that when a talker hears a new utterance, that utterance creates a distinct trace in short term memory. In addition, it activates existing memory traces that share similar features, such as previous iterations of the same lexical item. All traces combine, and their mean value becomes an *echo* in long term memory that drives a talker's standard production of the lexical item. The echo is dependent on the strength and content of the traces: traces with similar features that are present in higher numbers will influence a talker's production the most.

Exemplar theory explains the spontaneous phonetic convergence found in Goldinger (1998). Because an echo's strength and content is dependent on the number and similarity of traces in memory, during an auditory naming task infrequent and unfamiliar utterances should activate fewer existing memory traces, leaving the shadower more susceptible to the features of the target talker's utterance. In his phonetic convergence experiment, Goldinger found when the target talker utterance was a low frequency word, more phonetic convergence occurred. Additionally, under exemplar theory, increasing the number of times a shadower hears a target word provides more identical traces to memory, generating an echo that approximates the features of the repeated utterance. Accordingly, Goldinger also found that shadowers showed more phonetic convergence to words they had heard multiple times. He concluded that words that do not occur often result in less trace variation, while more repetitions of a particular utterance create an echo that is more similar to that utterance.

Combining social and automatic theories, Babel (2009) suggested that phonetic convergence is an exemplar-based process, but that it is mediated by social factors. In her study, she examined phonetic convergence of shadowers who repeated low frequency words containing

the vowels /i, æ, ɑ, o, u/ spoken by one black and one white target talker. As mentioned above, she found that shadowers who demonstrated a pro-black racial bias in an Implicit Association Task converged more to a black target talker than did shadowers with less of a pro-black bias, indicating that social factors play a role in determining phonetic convergence. However, results also revealed that only the vowels /æ/ and /ɑ/ underwent convergence, while /i/. /o/, and /u/ did not. Babel's explanation for this finding was that compared to /i/, /o/, and /u/, the low vowels /æ/ and /ɑ/ show much more possible variation in pronunciation across dialects of American English. Therefore, for these vowels, participants have more production variety in their phonetic inventory to approximate when converging to another talker. Therefore, Babel argued that unconscious social preference predicts when phonetic convergence occurs, but that the process is restricted by the shadower's phonetic repertoire.

Assessing Phonetic Convergence

Two primary methods have been used to assess the presence and degree of phonetic convergence: a perceptual AXB task and acoustic analysis. In a typical AXB paradigm, a listener hears three versions of the same word where X is the target talker's production, and A and B are two different productions by the shadower. One of these is the baseline production, where the words are read from a list, and the other is the shadower's imitation of the target talker's production from the auditory naming task. The listener's task is to decide which of the shadower's two productions, A or B, is a better imitation of the target talker's production, X. If the shadower did alter his/her production during the shadowing task, exhibiting phonetic convergence, then listeners should choose the shadower's shadowed productions over the baseline productions at a rate higher than chance.

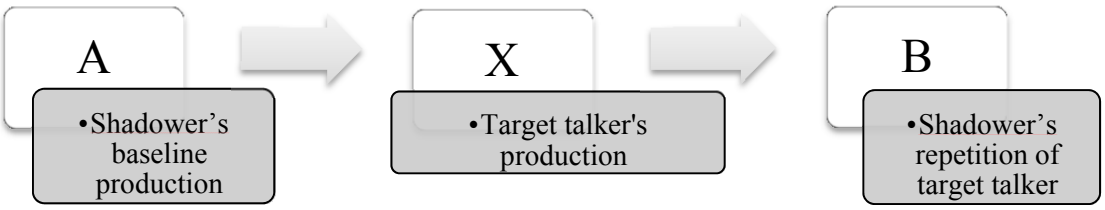

Figure 1. The AXB task, a perceptual measure of phonetic convergence

Because the listener's basis for selecting A or B is based on the entire word and not limited to a particular phonetic parameter, the AXB test allows for a holistic assessment of phonetic convergence. However, while this method allows listeners to use any or all phonetic parameters to judge convergence, it is impossible to determine which parameters were actually modified in the shadowed production.

The second method used to assess the presence and degree of phonetic convergence is with acoustic analysis. An advantage of acoustic analysis is that researchers are able to isolate and assess specific parameters that might be affected by phonetic convergence, such as f0 (Phillips & Clopper, 2012), VOT (Shockley et al., 2004), or vowel quality (Babel, 2009, 2012). Using acoustic analysis also allows researchers to detect changes in production that a listener might not perceive, while at the same time helps verify changes that listeners do perceive. However, a limitation of acoustic analysis is that it can only measure the contribution of preselected acoustic parameters to phonetic convergence. Since it is as yet unknown exactly which factors of an individual's speech affect phonetic convergence, acoustic analysis could produce misleading results; an analysis might indicate that no phonetic convergence occurred, when in fact it just did not manifest in whatever particular factor (e.g., f0, VOT) was measured.

The current study

To further investigate the effects of regional dialect variation and dialect distance on perceived phonetic convergence, the current study uses an auditory naming paradigm and AXB perception task. However, the current study differs from previous studies in two key aspects. First, the stimuli were specifically chosen to include distinct dialect differences between two regional dialects of English: New York City and General American. This manipulation was explored to help determine the effects of dialect-specific phonetic differences between the target talkers and shadowers on degree of phonetic convergence. Specifically, three distinct dialect differences were used between New York City versus General American: /ɔ/ vs. /a/ (e.g., 'caution'), /æ/ vs. /ɛ/ before intervocalic /r/ (e.g., 'barren'), and /ʃtr/ vs. /str/ (e.g., 'streamer') (see methods for more information).

Second, to test the differences between dialects, shadowers completed the full auditory naming task with two target talkers, one from each dialect, rather than producing a baseline and shadowed production. As a result of this methodological change, the AXB paradigm in the current study used two shadowed productions for the A and B stimuli.

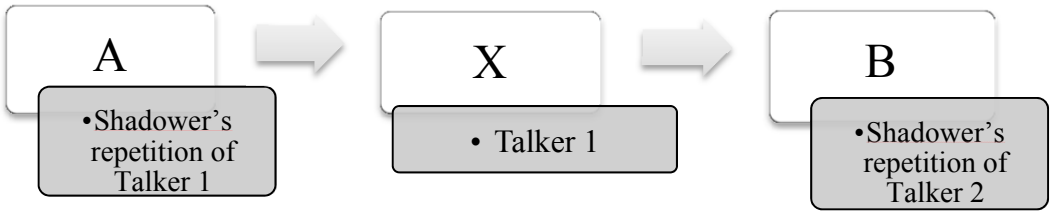

Figure 2. Sample AXB trial for the current study

Given the conflicting findings on whether more or less similarity between shadower and target talker produces more phonetic convergence, these differences were chosen to help shed

some light on the issue. First, if shadowers are sensitive to regional dialect variation within a word, then there should be a significant difference in degree of convergence between dialect-specific words and dialect-neutral words (words that have no known dialectal differences between New York City and General American). Specifically, the prediction is that shadowers will be perceived to converge more to the dialect-specific words due to the presence of the additional salient dialect-specific features to pick up on in the target utterances. Additionally, if dialect distance between shadower and target talker affects how much shadowers converge, it should be evident in two different effects. First, there should be a difference in the amount that the shadowers converge to the New York City target talker versus to the General American target talker. If shadowers converge when there is more similarity between them and the target talker as found in Kim et al. (2011), then the shadowers of the current study, none of whom speak with a New York City dialect, should converge more to the General American target talker, particularly in the dialect-specific condition. Conversely, if shadowers converge more when their dialect is *less* similar to the target talker as found in Babel (2012) and Walters et al. (2013), then there should be more convergence to the New York City target talker, particularly in the dialect-specific conditions. Finally, if dialect similarity does in fact predict degree of phonetic convergence, there should be no significant differences in convergence to the New York City versus General American talker in the dialect-neutral conditions, where there are no known dialectal differences in pronunciation between the New York City and General American dialects.

Therefore, the goals of the current study are (1) to replicate the findings of existing phonetic convergence studies that show spontaneous phonetic convergence in a shadowing paradigm, (2) to see if the degree of convergence differs in words that contain dialect-specific

contrasts versus words that show no dialect variation, and (3) to see if *dialect distance*, or the degree of similarity of regional dialects of shadower and target talker, has an effect on the degree of phonetic convergence.

Methods

Target Talkers

Two speakers were selected from a larger group to serve as the target talkers. Both talkers were female native speakers of American English with no reported speech or hearing problems. Talker 1 was 20 years-old and spoke with a local New York City dialect. Talker 2 was 23 at the time of recording and spoke with a General American English dialect.

The talkers recorded a set of 105 words, 36 of which were selected to be used in the present study. All words were low frequency, at less than 75 occurrences per million (Kucera & Francis, 1967). Eighteen were considered to be dialect-neutral, containing no known dialect-specific differences between the New York City and General American talkers. The remaining 18 were selected to fall into three dialect specific groups (New York City vs. General American, respectively): /ɔ/ vs. /a/ (e.g., 'caution') (Labov et al., 2006), /æ/ vs. /ɛ/ before intervocalic /r/ (e.g., 'barren') (Clopper, Pisoni, & de Jong, 2005; Labov et al., 2006), and /ʃtr/ vs. /str/ (e.g., 'streamer') (Kraljic, Brennan, & Samuel, 2008). The complete set of words can be found in the appendix.

The talkers were recorded in a sound-attenuated IAC booth in the Department of Communicative Sciences and Disorders at New York University. Speech samples were recorded using a Sennheiser HMD 280-XQ-2 combination headphones & microphone, and recordings were taken using a Fostex FR-2LE recorder at a 44.1 kHZ sampling rate. They were transferred

to a PC as .wav files and were then amplitude-normalized to 65 dBHL and down sampled to 22.05 kHz for acoustic analysis with Praat (Boersma & Weenink, 2008).

During the recording sessions, words were presented on a laptop PC running on Windows XP and Eprime 2.0 sitting outside the booth but visible through a large window. The words appeared on the screen for 1500 ms with 500 ms between each word. Each talker produced the list five times in a random order, and the best repetition was chosen for the experiment to avoid creakiness, coughs, or other vocal mistakes or interference.

Shadowing Talkers

Six shadowers were selected from a larger group to be used in the current study. All were female native speakers of General American English between the ages of 19-29 with no reported speech or hearing problems. All spoke with a General American dialect and used none of the variations typical of the regions in which some of them had lived (e.g., southern vowel shift (Labov et al., 2006)). Crucially, none of the shadowers had lived in the New York City area prior to age 18, and assessment of each of the shadowers' first repetitions of the New York City target talker revealed that none used the New York City dialectal variations /ɔ/, /æ/ before intervocalic /r/, and /ʃtr/. The dialectal proximity of Shadower 2 to the New York City region and its possible effects on her results will be expanded upon in the discussion section. Additional demographic information can be seen in Table 2.

	Shadower 1	Shadower 2	Shadower 3	Shadower 4	Shadower 5	Shadower 6
Age	19	20	29	19	19	20
Talker heard first	New York City	New York City	General American	General American	New York City	General American
Dialect	West	Hudson Valley, NY	Texas/DC	Georgia/DC	Midland	West

Table 2. Shadower information

For the phonetic convergence task, the shadowers completed an auditory naming task in which they were instructed to identify the word they heard over headphones by speaking it into the microphone. Shadowers completed two blocks separated by target talker and counterbalanced for talker order across participants (see Table 2). In each block, the 36 target words were presented three times in random order.

Words were presented over headphones with a 3000 ms interval between each word. Recordings were made in the same manner as the target talkers. The third repetition from the presentation of each target talker was selected to be used in the AXB listening task. This last repetition was selected both to avoid any hyperarticulation that might have occurred in the first repetition and to ensure that the shadowers had an optimal amount of exposure to the words to induce phonetic convergence (Goldinger, 1998). In a small number of cases (23/432), the second repetition was selected to be used due to mistakes in the third repetition. The selected words from each block were spliced out in Praat to be used as stimuli in the listening task.

AXB Listening Task

A new set of 96 native American-English participants between the ages of 18-29 (32 male) with no reported speech or hearing disorders was recruited to take part in the listening task. Separate AXB tasks were constructed for each of the six shadowers and given to 15-17 listeners each. For

each shadower, the listeners completed a perceptual AXB paradigm which was used to determine if a shadower's repetition of one target talker's production was more similar to that target than was the shadower's repetition of the other target talker. In each trial, listeners heard three versions of the same word, with one target talker's version as X and the shadower's two versions (one from the repetition of each target talker) of the same word as A and B. Each word triad was presented four times, (two target talkers X two orders of shadowed productions). The whole task was then repeated, resulting in eight presentations of each of the 36 words, for 288 trials total.

Listeners were told to decide whether the first or the third item (A or B) was a better imitation of the middle item (X), and they pressed the f (for first) or j (for third) keys on the keyboard to indicate their response on each trial. They were told to make their decisions as quickly and accurately as possible. The task was given over Sennheiser HD 280 PRO headphones in a quiet room using EPrime 2.0 on the same PC laptop used in the phonetic convergence task. The listening task took around 30-40 minutes to complete.

Results

Overall Pattern of Phonetic Convergence

The listeners' overall percent correct responses across all trials were averaged to determine whether listeners for a particular shadower were significantly above chance. A binomial test indicated that 53.4% was necessary to be considered significantly above chance. Listeners performed significantly above chance for five of the six shadowers (Shadower 1: 58.19%, Shadower 2: 53.57%, Shadower 3: 72.41%, Shadower 4: 56.56%, Shadower 5: 50.28%, and Shadower 6: 66.25%). Shadowers were also analyzed to assess the influence of dialect variation and target talker on phonetic convergence.

Dialect Variation and Distance

To assess the overall effects of dialect specificity (whether shadowers converged differently to dialect-specific words versus to dialect-neutral) and target talker (whether shadowers converged differently to the New York City and General American target talkers), a repeated-measures ANOVA was run with dialect specificity (neutral, specific) and talker (New York City, General American) as within-subjects factors and shadower as the between-subjects factor.

As mentioned in the introduction, if language distance affects phonetic convergence, then there should be a difference between the amount of phonetic convergence in the dialect-specific condition and the dialect neutral conditions. Specifically, if being more similar to the target talker results in greater convergence, as argued by Kim et al. (2011), then convergence is expected to be greater for the General American target talker, particularly in the dialect-specific condition, in which dialect-related differences between General American and New York City are more salient. If, in contrast, being less similar to the target talker results in greater phonetic convergence, as argued by Babel (2012) and Walters et al. (2013), then convergence should be greater for the New York City target talker, particularly in the dialect-specific condition. It is expected that there be no significant difference between convergence to the target talkers in the dialect-neutral condition.

Following ANOVA measurements, planned paired-samples t-tests were run to determine whether perceived shadowing was different for the two target talker conditions separately for the dialect-neutral and dialect-specific conditions. The results are summarized in Figure 3.

Results revealed a significant main effect of talker ($F(1,90) = 37.536$, $p<.001$), with greater shadowing for the General American talker (62%) than the New York City talker (57%).

A significant main effect of dialect specificity was also found ($F(1,90) = 13.936$, $p<.001$), with more shadowing in the dialect-specific condition (61%) than in the dialect-neutral condition (58%). A significant main effect of shadower was also found ($F(1,90) = 19.248$, $p<.001$). Significant interactions were found between target talker and shadower ($F(5,90) = 4.490$, $p=.001$) and between dialect specificity and shadower ($F(5,90) = 2.486$, $p=.037$). Interaction between target talker and dialect specificity was not significant ($p=.694$). Finally, a three-way interaction between target talker, dialect specificity, and shadower was found ($F(5,90) = 3.159$, $p=.011$).

Because results revealed a significant main effect of shadower, a significant interaction between shadower and target talker, and a significant interaction between shadower and dialect specificity, separate repeated-measures ANOVAs were run on each of the five shadowers with perceived shadowing significantly above chance with dialect specificity (neutral, specific) and talker (New York City, General American) as within-subjects factors.

For Shadower 1, neither main effect nor the interaction reached significance, although the main effect of target talker approached significance ($F(1,16) = 3.555$, $p=.078$). The difference in convergence between the two target talkers in the dialect-neutral condition approached significance ($t(16)= -1.892$, $p=.077$) but no significant difference was found for the dialect-specific condition ($p=.443$).

Shadower 2 did not show a significant main effect of either dialect specificity or target talker ($p>.30$), but the interaction approached significance ($p=.095$). The difference in convergence in the dialect-neutral condition approached significance ($t(16)= -1.780$, $p=.094$), and but no significant difference was found in the dialect-specific condition ($t(16)=1.019$, $p=.323$).

Shadower 3 showed a significant main effect of dialect ($F(1,14) = 5.048$, $p=.041$) and a significant main effect of target talker ($F(1,14) = 41.434$, $p<.001$). An interaction between dialect and target talker approached significance ($p=.088$). In addition, Shadower 3 showed significantly more convergence toward the General American target talker than the New York City target talker in the dialect-neutral condition ($t(14)= -4.107$, $p=.001$) and also in the dialect-specific condition ($t(14)= -5.835$, $p<.001$).

For shadower 4, no significant main effect of dialect was found ($p=.318$), but there was a significant main effect of target talker ($F(1,15) = 21.748$, $p<.001$) and a significant interaction between dialect and target talker ($F(1,15) = 5.754$, $p=.030$). Like Shadower 3, Shadower 4 also showed significantly more convergence toward the General American target talker than the New York City target talker in both the dialect-neutral ($t(15)= -2.20$, $p=.044$).condition and dialect-specific condition($t(15)= -4.65$, $p<.001$).

Finally, for shadower 6, results revealed a significant main effect of dialect ($F(1,15) = 16.019$, $p=.001$) and a significant main effect of talker ($F(1,15) = 6.779$, $p=.020$). The interaction between dialect and target talker did not reach significance ($p=.169$). In addition, no significant difference was found between the two target talkers in the dialect-neutral condition ($p=.106$), but results showed significantly more convergence toward the General American target talker than the New York City target talker in the dialect-specific condition ($t(15)= -3.25$, $p=.005$).

Finally, because greater convergence was found in the dialect-specific conditions, the distribution of convergence to the dialect-specific variations (/ɔ/ vs. /a, /æ/ vs. /ɛ/ before intervocalic /r/, and /ʃtr/ vs. /str) was analyzed per shadower. No clear patterns emerged in overall convergence to the dialect-specific variations. The distribution can be seen in Figure 4.

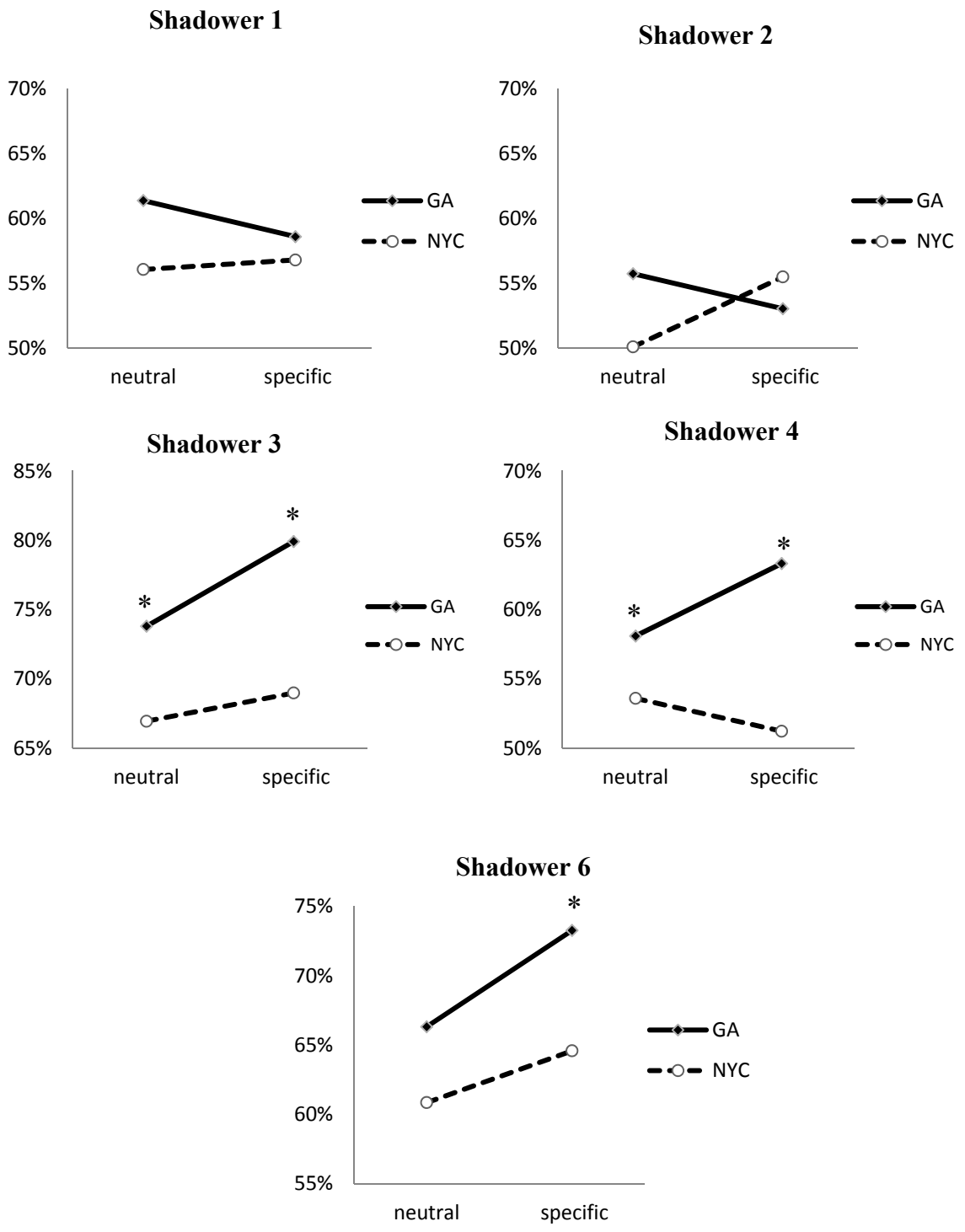

Figure 3. Differences in convergence to target talkers per dialect specificity condition. A * indicates a significant difference.

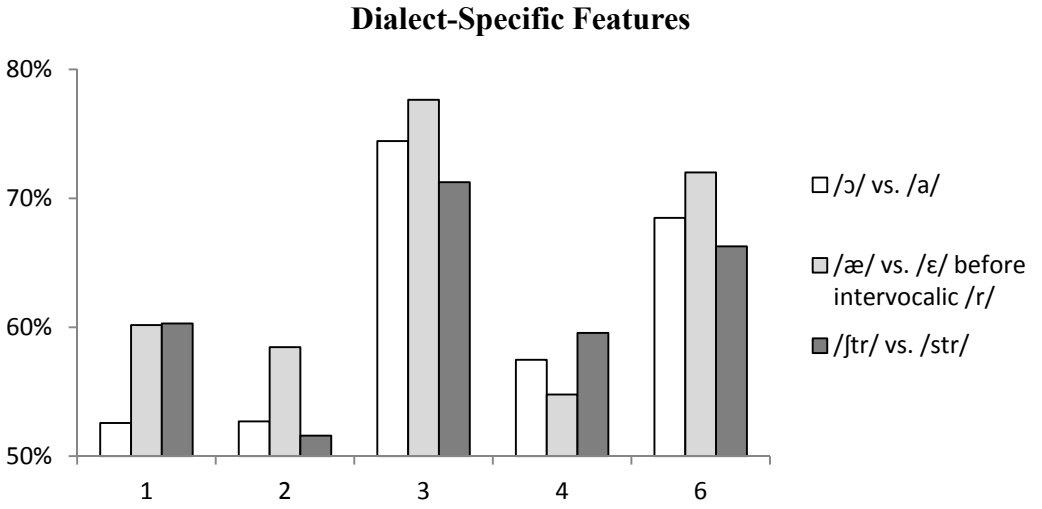

Figure 4. Convergence per shadower to dialect-specific features

Summary of Results

To summarize, five of the six shadowers converged to the target talkers. Overall, shadowers converged more to the General American target talker than the New York City target talker. In addition, shadowers overall converged more to the dialect-specific words than the dialect-neutral words. Individual shadowers showed varied patterns of convergence. Of the five shadowers who converged at a rate higher than chance, two showed significantly greater convergence in the dialect-specific condition, while the other three did not show significant effects of dialect specificity. Three shadowers converged to the General American target talker at a rate above chance, while the other two did show significant effects of target talker. There was no clear pattern across shadowers of convergence to the dialect-specific variations. Finally, three shadowers showed significantly more convergence to the General American target talker than to the New York City target talker in the dialect-specific condition, but two of those shadowers also showed significantly more convergence to the General American target talker in the dialect-neutral condition.

Discussion

Overall convergence

Results of the AXB listening task revealed that five of the six shadowers showed significant phonetic convergence overall, confirming the results of prior studies that found spontaneous phonetic convergence in a shadowing paradigm. The failure of Shadower 5 to either converge to or diverge from the target talkers as well as the relatively wide range in accuracy scores of the other shadowers (from 53.5% to 72.4%) and the variation in findings of studies discussed in the introduction indicate that phonetic convergence is dependent on factors specific to individual shadowers and is not a fully automatic process.

Language Distance

The clearest pattern that emerged from the current study was a significant effect of target talker across three shadowers (3, 4, and 6) and an effect that approached significance for a fourth shadower (1), which indicates that language distance between the shadower and target talker influences phonetic convergence. Specifically, shadowers who exhibited an effect of target talker were all more likely to converge to the General American Talker. As discussed in the introduction, it is unclear in previous research whether more or less similarity between the dialects of the shadower and target talker facilitates phonetic convergence. The results of the current study generally support the finding of Kim et al. (2011) that shadowers are more likely to converge to a talker with a similar dialect. All conditions of greater convergence to a particular target talker were those in which there was less dialect distance between target talker and shadower.

If shadowers are more likely to converge to target talkers who sound more like they do, there are two implications for what the mechanisms are that underlie phonetic convergence. One theoretical explanation for the finding, put forth by Kim et al. (2011) and applicable to the results of the current study, is consistent with the Babel (2009) argument that phonetic convergence, while a semi-automatic, socially-motivated process, is limited to the phonetic constraints of the shadower's repertoire. Therefore, a shadower whose dialect is considerably dissimilar from the dialect of the target talker would not converge if the target talker's utterances include sounds outside the shadower's repertoire (e.g., for a General American shadower, /ɔ/, /æ/ before intervocalic /r/, and /ʃtr/). At the same time, a shadower and target talker with less language distance between them are more likely to have similar phonetic repertoires, allowing more opportunities for phonetic convergence.

Second, the methodology of the current study weakens the argument that phonetic convergence to dialect is an overtly socially-motivated process to manage social identity and relationships, in accordance with Communication Accommodation Theory (Giles, 1973). The current study used a socially-minimal auditory naming task to elicit convergence, suggesting that the effect of language distance was not likely to be motivated, for example, by a conscious desire of the shadowers to maintain social distance between themselves and the New York City target talker or minimize differences between themselves and the General American talker. While it could still be the case that subtle biases do come into play (for example, the Adank et al. (2013) study on imitation of single words spoken in Glaswegian English found that participants were nevertheless able to rate the target talker on multiple social parameters and that their ratings were affected by imitation condition), it appears from the current study that a more automatic process governs the effects of dialect difference on phonetic convergence. Kim et al. (2011) used a

socially-rich conversation-elicitation task to bring about convergence, making the current study the first socially-minimal study to suggest that shadowers are more likely to converge to a target talker who sounds more like they do.

Dialect-Specific Variations

The results of the remaining factors investigated in this study were less clear. No clear pattern emerged across the shadowers with respect to the degree of convergence in the three dialect variations, /ɔ/ vs. /a/ (e.g., 'caution'), /æ/ vs. /ɛ/ before intervocalic /r/ (e.g., 'barren'), and /ʃtr/ vs. /str/ (e.g., 'streamer'). This indicates two possibilities. The first, also suggested by the varied degrees of overall convergence found, is that there is a large amount of individual variation in the factors that influence shadowing, and that factors influence each shadower differently. In the current study specifically, variation among shadowers in the different factors affecting phonetic convergence is likely. First, although none of the shadowers came from the New York City dialect region, therefore sharing a lack of exposure to the dialect, they all came from different regions (West, Midland, South, and Hudson Valley, NY) and therefore had considerably different language experiences. In addition, because the study was conducted in New York City, all shadowers were living here at the time of the experiment, and length of residence in New York City at the time of recording varied from six months to three years, meaning that all shadowers had been exposed to different amounts of the New York City dialect. How much experience with a dialect is needed to change an adult's perceptions and, ultimately, productions? Previous research has shown that both temporal and social factors affect dialect change. Kraljic et al. (2008) investigated the effects of short-term dialect training by comparing the effects of acoustic variation caused by regional dialect to effects of the same variation caused

by idiolect on listeners' phoneme boundaries. They used the /ʃ/ realization of /s/ typical in the /str/ cluster of the New York City dialect in two contexts: one condition exposed participants to the /ʃ/ variant in dialect-independent contexts such as *pedestal*, and the other presented the variant in dialect-specific contexts such as *district*. Following exposure, they measured the /s/-/ʃ/ perceptual boundary for each participant and found that while the participants in the dialect-independent condition changed their perceptual boundaries, participants in the dialect-dependent condition did not, indicating that perceptual learning does not immediately occur upon introduction to a dialect-specific variation. However, it has been found that talkers new to a region begin to undergo long-term dialect change as early as three months after exposure to the new dialect (Evans & Iverson, 2007). What is more, in addition to temporal factors, social factors can also influence whether a dialect is learned. Under Communication Accommodation Theory, for example, it can be predicted that speakers who desire to be accepted by a new community will be more likely to take on the characteristics of the community's dialect, while speakers who wish to maintain their social identity or distance themselves from the new community will be less likely to change their dialect (Giles, 1973).

A second possibility for the variation in perceived convergence to the dialect-specific condition has to do with the backgrounds of the listeners in the AXB task. As discussed in the introduction, phonetic convergence judgements in the AXB task depend entirely on the listeners' abilities to perceive phonetic change. A listener might not be able to pick up certain changes that would be evident in an acoustic analysis such as, for example, if a General American shadower hearing the word *streamer* pronounced by a New York City target talker as /ʃtrimɚ/ slightly modifies her production of the word from the General American variant /strimɚ/ to an ambiguous /s~ʃtimɚ/.

A listener's ability to pick up on the phonetic changes a shadower may have made depends on multiple factors. First, studies on perceptual categorization of regional dialect have found that geographic mobility and location of origin affect listeners' dialect categorization accuracy as well as their perception of dialect variation: generally, listeners who have lived in more than one region are more accurate at perceptual dialect tests, and listeners originating from a particular region are more accurate at perceiving and categorizing the dialect of that region than non-residents (Clopper & Pisoni, 2004, 2006; Evans & Iverson, 2004). The implications of these findings are evident in the following example using the sample AXB trial,

$$/bærən/_{\text{shadowed NYC}}\ ,\ /bærən/_{\text{target NYC}}\ ,\ /bɛrən/_{\text{shadowed GA}}$$

in which the shadower converges to the New York City target talker's /æ/. A non-mobile General American listener who has the merry-Mary-marry merger of /ɛ/, /eɪ/, and /æ/ before intervocalic /r/ (Labov et al., 2006) might not detect the difference between /bærən/ and /bɛrən/, whereas a New York City listener, for whom all three vowels are distinct (Labov et al., 2006), might. It would be expected that the General American listener would not pick up on the altered vowel and instead be forced to reply solely upon other factors that might have changed during convergence (e.g., speaking rate, F0, F0 trajectory) to make his decision. The New York City listener, on the other hand, would be able to use the vowel change as an additional salient cue to phonetic convergence and might perform better than the General American listener as a result. A similar advantage for the New York City listener over listeners from other dialect regions could be expected with the /ɔ/ vs. /a/ (e.g., 'caution') variation in the current study due to the merger of those two vowels in over half the of the North American region, including the West, Eastern New England and Western Pennsylvania (Labov et al., 2006).

Interestingly, the New York City could be at a disadvantage relative to other listeners when faced with the /ʃtr/ vs. /str/ contrast. To the General American Listener, the contrast between /ʃ/ and /s/ in the sample AXB trial,

$$/ʃtrimɚ/_{\text{shadowed NYC,}}\ /ʃtrimɚ/_{\text{target NYC}}\ ,\ /strimɚ/_{\text{shadowed GA}}$$

would be highly salient due to the /ʃtr/ cluster's existence primarily in Philadelphia and New York City English (Kraljic et al., 2008), enabling the General American listener to use it to help make a decision. In contrast, since the New York City dialect contains the cluster, the difference between the shadowed tokens in the AXB trial might not be as obvious to the New York City speaker.

On the other hand, the /ʃtr/ vs. /str/ variation could also present a problem for a listener whose dialect does *not* include the /ʃtr/ cluster in a situation such as the sample AXB trial,

$$/strimɚ/_{\text{shadowed NYC,}}\ /ʃtrimɚ/_{\text{target NYC}}\ ,\ /strimɚ/_{\text{shadowed GA}}$$

in which the shadower does not converge to the dialect-specific variation. In this situation, it is possible for the /ʃtr/ in the X position to be so salient and out of the ordinary to the listener that neither A nor B would seem like a better imitation of X, leading the listener to miss other factors the shadower might have changed during convergence and perform at chance for the trial. Similarly, a New York City listener could be put at a similar disadvantage when presented with AXB trials in which the shadower does not pick up on the New York City /ɔ/ or /æ/. The absence of the salient dialect markers from the A and B tokens could obscure other altered characteristics.

It is clear that the language backgrounds of the listeners have the potential to significantly affect the results of an AXB test. To investigate whether the dialect and residence history of listeners affects their performance on an AXB test, a study controlling for listener dialect background should be conducted. In addition , it will be necessary in a future experiment to

compare the perceptual results from the current study to results from an acoustic analysis. In fact, Pardo et al. (2013) argued the importance of employing both perceptual and acoustic measurement methods in phonetic convergence studies following an experiment in which she directly compared the findings of acoustic analysis and of AXB listening tests and found that none of the patterns found in acoustic measurements agreed with the perceptual data. Phillips and Clopper (2012) also found discrepancy between the acoustic features that acoustic analysis found to have converged and the features that correlated with the AXB results. Acoustic analysis showed significant convergence only in coda duration, while both coda duration and vowel quality correlated with the AXB results. For the current study, acoustic analysis will also be helpful to further investigate the patterns observed for the significant differences in convergence to the two target talkers, which contrary to the predictions made, were also found in the dialect-neutral condition for two shadowers.

Additional Possible Causes of Variation in the Current Study

The following paragraphs will explore additional linguistic and methodological factors that might have contributed to the variation found among the shadowers of the current study. First, dialect and region of origin are not the only factors that might make for better listeners. For example, advanced musical training has been linked to improvements in speech and language processing. Citing several recent studies, Patel (2014) posited that repeated, attention-demanding musical training that (1) uses brain networks overlapping with those associated with speech processing and (2) places higher demand on that network than speech processing, can benefit processing of shared acoustic features such as pitch, amplitude, and rhythm. Because the current study used a single-word auditory naming paradigm, there would little chance for convergence

based on rhythm; F0 trajectory, however, has been shown to be affected by convergence (Phillips & Clopper, 2012). Additionally, Chobert, Francois, Velay, and Besson (2014) found that 8-10 year-old children showed enhanced processing of syllable duration and VOT after 12 months of music training compared to age-matched children who received 12 months of painting training. Both syllable duration and VOT have been shown to be affected by phonetic convergence (Phillips & Clopper, 2012; Shockley et al., 2004). Consequently, it could be hypothesized that listeners with advanced musical training would be more accurate at the AXB task. Participants in the current study filled out a questionnaire which asked them to list, along with language and residence background, any musical training they had received. Several listeners reported over 10 years of either vocal or instrumental training. While there were too few musically-trained listeners in each shadower condition and too much variation in type of training to analyze in the current study, either a future study that controls for listener experience or one that includes an experimental musical training period could expand the research on musical training and linguistic perception to include the effects of musical training on phonetic convergence.

A possible methodological cause of the variation in the current study has to do with the order of presentation of the target talkers to the shadowers. Interestingly, all three shadowers who converged significantly more to the General American target talker than to the New York City target talker (3, 4, and 6) heard the General American target talker first. The three shadowers who heard the New York City target talker first either did not converge at all (5) or showed no significant effects of dialect or target talker (1 and 2). This could indicate a possible task effect on convergence to different target talkers which could be explored in a future study.

Finally, returning to possible shadower-specific reasons for variation, it is important to note that Shadower 2, the only shadower who showed phonetic convergence but did not also

show an effect of target talker either approaching or reaching significance grew up in the Hudson Valley, a dialect region that, although historically distinct from the New York City region, has experienced diffusion of some features of the New York City dialect (Labov, 2007). Although her first repetition of the New York City condition in the auditory naming task was not found to contain the New York City dialect-specific variations, it is likely that Shadower 2 had been exposed to the New York City since childhood. The theory that dialect distance influences phonetic convergence would predict that she would therefore be less likely to converge to at a higher rate to one particular talker.

Conclusion

The data from this study have further evidenced the existence of spontaneous phonetic convergence in socially-minimal settings. In addition, they suggest that the dialect backgrounds and experiences of talkers play an important role in determining patterns of phonetic convergence. Namely, the study found that shadowers are more likely to converge to target talkers when there is less dialect distance between them. Further study with a more homogenized set of shadowers, controlled variation in characteristics of the AXB listeners, and acoustic analysis for comparison to the AXB results is necessary to shed more light on the more detailed factors at play and begin to piece apart the theoretical implications involved.

References

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychol Sci, 21*(12), 1903-1909. doi: 10.1177/0956797610389192

Adank, P., Stewart, A. J., Connell, L., & Wood, J. (2013). Accent imitation positively affects language attitudes. *Frontiers in Psychology, 4*, 280. doi: 10.3389/fpsyg.2013.00280

Babel, M. (2009). *Phonetic and social selectivity in speech accommodation.* (Doctor of Philosophy Dissertation), University of California, Berkeley.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society, 39*(4), 437-456. doi: 10.1017/S0047404510000400

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

Babel, M., McGuire, G., Walters, S., & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation *Lab Phonol* (Vol. 5, pp. 123–150).

Boersma, P., & Weenink, D. (2008). Praat: doing phonetics by computer (Version Version 5.0.40). Retrieved from http://www.praat.org/

Chobert, J., Francois, C., Velay, J. L., & Besson, M. (2014). Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cerebral Cortex, 24*(4), 956-967. doi: 10.1093/cercor/bhs377

Clopper, C. G., & Pisoni, D. B. (2004). Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change, 16*(1), 31-48. doi: 10.1017/S0954394504161036

Clopper, C. G., & Pisoni, D. B. (2006). Effects of region of origin and geographic mobility on perceptual dialect categorization. *Language Variation and Change, 18*(2), 193-221. doi: 10.1017/S0954394506060091

Clopper, C. G., Pisoni, D. B., & 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*(3 Pt 1), 1661-1676. doi: 10.1121/1.2000774

Evans, B. G., & Iverson, P. (2004). Vowel normalization for accent: an investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America, 115*(1), 352-361. doi: 10.1121/1.1635413

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: a study of accent change in young adults. *Journal of the Acoustical Society of America, 121*(6), 3814-3826. doi: 10.1121/1.2722209

Giles, H. (1973). Accent Mobility: A Model and Some Data. *Anthropological Linguistics, 15*(2), 87-105. doi: 10.2307/30029508

Giles, H., Mulak, A., Bradac, J., & Johnson, P. (1987). Speech accommodation theory: The first decade and beyond *Communication Yearbook 10* (pp. 13-48). Beverly Hills, CA: Sage.

Goldinger, S. D. (1998). Echoes of Echoes? An Episodic Theory of Lexical Access *Psychological Review, 105*(2), 251-279. doi: 10.1.1.380.9874

Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic bulletin & review, 11*(4), 716-722. doi: 10.1.1.381.2311

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

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: dialects, idiolects, and speech processing. *Cognition, 107*(1), 54-81. doi: 10.1016/j.cognition.2007.07.013

Kucera, H., & Francis, W. N. (1967). *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press.

Labov, W. (2007). Transmission and diffusion. *Language*, 344-387. doi: 10.1.1.189.1124

Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin: Mouton de Gruyter.

Miller, R. M., Sanchez, K., & Rosenblum, L. D. (2010). Alignment to visual speech information. *Attention, Perception, & Psychophysics, 72*(6), 1614-1625. doi: 10.3758/APP.72.6.1614

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: the role of perception. *Journal of Language and Social Psychology, 21*(4), 422-432. doi: 10.1177/026192702237958

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology, 32*(5), 790-804. doi: 10.1037/0022-3514.32.5.790

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119*(4), 2382-2393. doi: 10.1121/1.2178720

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics, 40*(1), 190-197. doi: 10.1016/j.wocn.2011.10.001

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics, 72*(8), 2254-2264. doi: 10.3758/APP .72.8.2254

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

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research, 308*, 98-108. doi: 10.1016/j.heares.2013.08.011

Phillips, S., & Clopper, C. G. (2012). Perceived imitation of regional dialects. *Proceedings of Meetings on Acoustics, 12*(060002). doi: 10.1121/1.4704668

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics, 66*(3), 422-429. doi: 10.3758/BF03194890

Walters, S. A., Babel, M., & McGuire, G. (2013). The role of voice similarity in accommodation. *Proceedings of Meetings on Acoustics, 19*(1), 060047. doi: 10.1121/1.4800716

Appendix - Complete list of shadowed words

Dialect-neutral	/ɔ/ vs. /a/	/æ/ vs. /ɛ/ before intervocalic /r/	/ʃtr/ vs. /str/
beacon	cauldron	barrel	straggle
brazen	caution	barren	stranger
comet	gaudy	carriage	stratum
dismal	sausage	garish	streamer
falcon	tawdry	parrot	stricken
feeble	tawny	tariff	stroller
gallop			
gusto			
kidney			
label			
malice			
maple			
pallet			
puzzle			
rustic			
staple			
stigma			
villain			