2012

Spontaneous Phonetic Imitation Across Regional Dialects

Ellen E. Dossey
Macalester College, ee.dossey@gmail.com

Spontaneous Phonetic Imitation Across Regional Dialects

Ellen E. Dossey

Macalester College

Honors Project in Linguistics
Advised by: Christina Esposito
Department of Linguistics
Macalester College
Spring 2012

Abstract

Speakers subconsciously alter their pronunciation towards that of their conversation partner through a process called spontaneous phonetic imitation. Previous research has shown that this imitation may not be a completely automatic process. One factor that may influence the extent of imitation is social biases based on gender, age, or region. The current study used a lexical shadowing task to determine if speakers would spontaneously imitate American English vowel variants that were not found in their native dialect. Results showed that in the non-interactive task, regionally distinct vowels were significantly imitated. Furthermore, the level of imitation was not related to previous familiarity or experience with the dialect in question.

Spontaneous Phonetic Imitation Across Regional Dialects

An individual's pronunciation is characterized by the way they employ different sounds, stress, intonation, and voice quality. The typical pattern of pronunciation associated with a dialect is known as an "accent," and it is commonly acknowledged that this can be one of the most obvious markers of a speaker's regional or social routes. However, in some situations one's accent might not *always* be a reliable indicator of their regional routes; anecdotally is not uncommon to hear claims of people "losing," "changing," or "neutralizing" their accents throughout childhood or adulthood (Trudgill, 1986). Linguists have found evidence for these types of long-term changes, both in terms of speech production and perception. For example, Munro et al. (1999) found perceptual evidence of accent change in Canadians who had moved to Alabama as adults. The investigators assessed the dialectal change by having native speakers of Canadian or Alabaman English listen to recordings of speakers in three dialect groups: Canadians living in Canada, Canadians living in Alabama, and Alabamans living in Alabama. The listeners rated accents holistically on a scale of 1-9, with 1 corresponding to "more Canadian," and 9 corresponding to "more American." The mean ratings of each accent-type proved to be significantly different, suggesting that the Canadians living in Alabama sounded more American than the Canadians living in Canada; however, they also still spoke distinctly from their Alabaman counterparts, meaning that they had developed an intermediate American-Canadian accent. A second experiment by Munro et al. (1999) examined this question on a segmental level; listeners were presented with six tokens containing the vowel /a͡ɪ/, which is of interest because of its distinct pronunciations in both Canadian English [ʌ͡ɪ] and Alabaman English [a]. Again, ratings were given on a scale of "more Canadian" to "more American," and it

was found that accent change may be detectable over single-word utterances if the words contain a segment that differentiates the two accents.

Evans and Iverson (2007) further explored segmentally-based accent change. They quantifiably examined the speech of London college students who were originally from the north of England. The Northern English (NE) dialect spoken by the students upon their arrival to university contained many vowel segments that were not characteristic of Southern Standard British English (SSBE), which is commonly spoken in London. The study focused on these segments, specifically examining the vowels in words like "bud" and "bath," respectively. In NE, the "bud" vowel is pronounced as a high back variant, /ʊ/, and in SSBE it is more centralized. The word "bath" in NE contains the vowel /a/, whereas in SSBE it contains the vowel /ɑ/. Over a period of two years, acoustic analysis of these sounds in the students' speech often showed a significant change towards the SSBE variants (Evans & Iverson, 2007).

Studies such as those described above support the notion that individuals' accents can change over time. The next step in this discussion is to explore the question of *why* these types of changes occur, what cognitive and linguistic processes underlie accent shifts, and what social situations are necessary for these processes to take place. Linguistic theory hypothesizes that phonetic imitation, the process through which children acquire native pronunciations of their mother tongue, never stops; even in adulthood we are listening to others' speech and ever-so-subtly adjusting our own pronunciation (Babel, 2009). The extent to which we do this may be dependent on a number of other factors. In a discussion of regional accent change, one factor that has to be taken into account is sociolinguistic variables. Social psychologists and linguists have written extensively on situational-based speech accommodation (see Labov, 1994; Giles, 1975; Trudgill, 1986). This paper will draw from both disciplines to explore, experimentally, the

possibility for spontaneous phonetic imitation across dialects, and to look at the possible effect that previous experience and social stigma have on one's likelihood to accommodate to a regional accent in a shadowing paradigm. Imitation will be analyzed at the level of individual sounds, specifically looking at vowels. For this reason the first part of this paper will be a brief review of the phonetic characteristics of vowels, followed by a review of previous research concerning the specific cognitive and social factors that affect both short-term and, by extension, long-term intraspeaker language variation and change.

Acoustic features of accent variation: a look at vowels

Most speech sounds are created when air flows from the lungs through the larynx and exits the mouth. This path is known as the vocal tract (see Figure 1). Sounds are differentiated by modifications of this airstream as it flows through the vocal tract; these modifications are based on vibrations of the vocal folds (vocal chords), constriction of the vocal tract, and articulator placement and behavior.

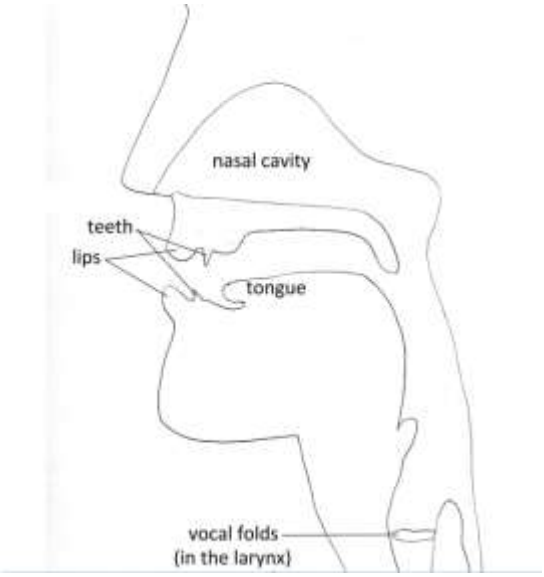
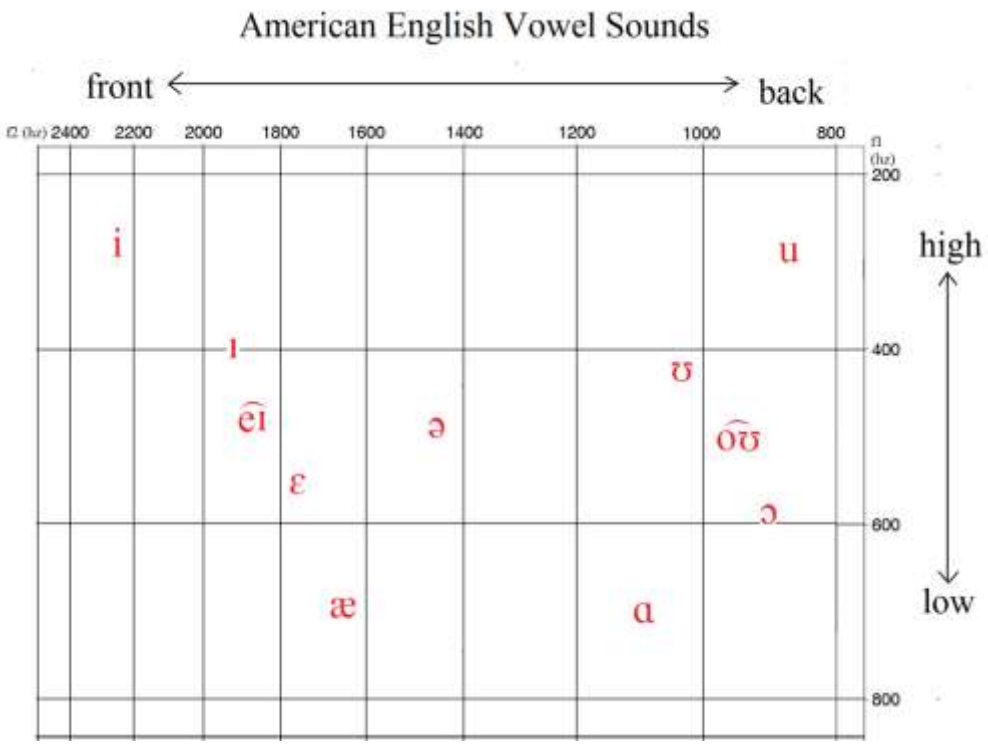


Figure 1: The human vocal tract.

In the case of consonants, there is some sort of closure or constriction in the vocal tract as sounds are being pronounced. For example, if you pronounce the sound represented by the letter "t," you can feel the front of your tongue touching the ridge behind your top teeth, causing a momentary blockage of the airstream before it is released and the sound is produced. Though different consonants feature different amounts of closure at different locations within the vocal tract, they all block airflow in a specific way, usually using the tongue.

Vowels, on the other hand, are the sounds that are produced when the vocal tract is open and air flows freely. They are differentiated by three factors: rounding, tongue height, and tongue backness. Rounding refers to whether or not the lips are rounded during pronunciation; height and backness refer to the position of the tongue in the mouth relative to the top of the mouth and back of the mouth. Varying the tongue position does not create any closure, but does affect the shape of the vocal tract, which creates different vowel sounds. Vowel sounds are produced on a spectrum, from highest to lowest and from farthest back to farthest fronted (see Figure 2 for a chart illustrating the vowels of American English)

The fact that vowels exist on a spectrum means that vowels pronunciations are extremely variable, which is important when one considers variation in speech. At the level of the individual, pronunciation varies to a certain extent as the tongue approximates the specific location of the sound. For example, an individual's /i/ might be slightly farther forward or slightly higher (or vice-versa) from one pronunciation to the next. Generally, however, the relationship of vowels remains constant; for example, a person's /i/ is always higher than their /ɪ/ which is higher than their /ɛ/, etc.



Example words for each vowel

i	“leak”	u	“Luke”		
ɪ	“lick”	ʊ	“look”		
eɪ	“lake”	oʊ	“bloke”		
ɛ	“leg”	ʌ	“luck”	ɔ	“law”
æ	“lack”	ɑ	“lock”		

Figure 2. Chart of the American English vowel sounds (formant values taken from Ladefoged, 1975). The vertical axis represents the height of the vowel, and the horizontal axis represents the backness of the vowel. The table below the chart gives example words for each vowel. Additional notes: English also has three other diphthongs: /aɪ / (as in “like”) /aʊ/ (as in “cow”) and /ɔɪ/ as in “boy.” Also, the vowels /ɔ/ and /ɑ/ have merged for many speakers of American English, meaning that the word “law” is pronounced the same as the word “la” (both are /lɑ/)

Vowel sounds are measured in terms of their acoustic properties, specifically their formant structure. In phonetics, the term "formant" refers to areas of acoustic resonance in the vocal tract. When the vocal folds open and close, there is a pulse of air from the lungs which disturbs the air in the vocal tract, setting it into vibration. In a vowel sound, the air in the vocal tract vibrates at a number of different frequencies simultaneously. The pattern of vibrations varies depending on the shape of the vocal tract. Because all vowels are distinguished by slight differences in the shape of the vocal tract depending on tongue placement, they also have their own distinct formant structures which can be used to distinguish them from one another (Ladefoged, 1975.)

Formant structure can easily be analyzed using a spectrogram; in spectrogram images, the formants appear as dark bars. Though vowels have at least three visible formants in a spectrogram, the first two are generally sufficient for vowel identification. The first formant (F1) represents the vowel height; higher vowels have lower F1 measurements (in Hertz). The second formant (F2) is a measure of vowel backness—a higher F2 means that the vowel is farther back. Figure 3 provides spectrograms representing the vowels /ɑ/ and /i/. In these images, you can see that because /ɑ/ is a low back vowel, it has a high F1 and a low F2; the formants are practically touching. On the other hand, the high, front vowel /i/ has a low F1 and a high F2. There is a huge gap between the first two formants.

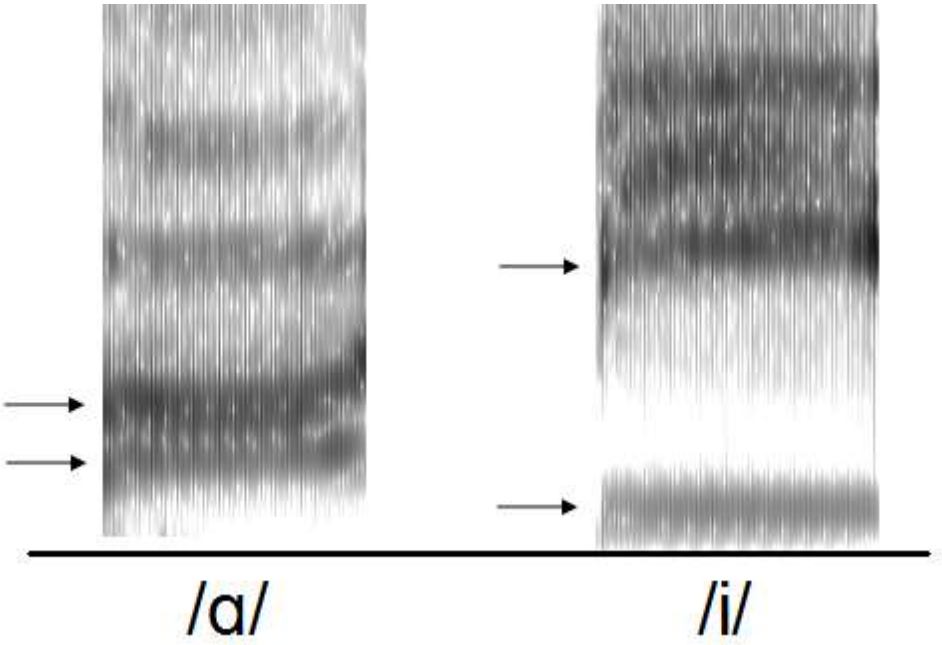


*Figure 3.*Spectrograms of the vowels [ɑ] and [i]. The first two formants are marked with arrows.

It is entirely possible, and actually quite common, that one person's vowels overlap in frequency with *different* vowels in another person's speech; for example, one person's /i/ might have the same acoustic measurements, on average, as another person's /ɪ/. This variation is caused by differences in the shape and size of the vocal tracts of different people and by regional and social trends in pronunciation.

The fact that vowels demonstrate so much easily-measurable variation makes them the perfect variable to focus on in studies of phonetic imitation. Participants in this study will be exposed to an accent whose vowels have a significantly different formant structure from the participant's own accent. We will study imitation by measuring changes in formant structure after participants have been exposed to the experimental stimuli.

Cognitive factors in accent change: phonetic imitation

A possible cognitive theory that explains intraspeaker language change as a result of phonetic imitation is based on the premise that every time we are presented with a new stimulus,

it is encoded as a unique memory trace; these memory traces are grouped into categories based on similarities with other previously encountered stimuli. This represents an episodic, or exemplar, theory of memory (Goldinger, 1997). When a new stimulus is encountered, all past traces, or exemplars, of the same category are activated in proportion to their similarity to the new stimulus. The traces that are most activated then come to consciousness and allow us to recognize the stimulus (Goldinger, 1997). In terms of spoken words, exemplar theory can be applied to both lexical and segmental memory (Goldinger, 1998; Babel, 2009). Unlike other theories of spoken-word perception and memory, which rely on the concept of normalization and therefore suggest that our memories ignore "irrelevant" information like speaker idiosyncrasies in order to activate a stored, abstract notion of the stimulus, exemplar theory allows for idiosyncratic representations to be stored and *composited* to form a general representation of a stimulus when the various memory traces are activated (Goldinger, 1997). Pierrehumbert (2000) explains the concept in the following way:

> Each category is represented in memory by a large cloud of remembered tokens of that category. These memories are organized in a cognitive map, so that memories of highly similar instances are close to each other and memories of dissimilar instances are far apart. The remembered tokens display the range of variation that is exhibited in the physical manifestations of the category.

Therefore when we are listening to speech, we are able to recognize unfamiliar voices saying familiar words, and at the same time we are able remember and identify specific, familiar voices with ease, even after significant periods of time have elapsed (Johnson, 1996).

It is important to clarify that our brains do not necessarily store a separate exemplar for every perceptual experience. If an encountered stimulus is identical to an already-stored exemplar, then it is classified as a repetition of that exemplar and a new trace is not formed; instead, the previous exemplar is strengthened. Also, because memories fade over time, recently

encountered exemplars have more relative weight than exemplars that have not been encountered for some time. Because of this, exemplars which represent frequent and/or recent experiences have higher resting activation levels than exemplars corresponding to infrequent and/or non-recent experiences. (Pierrehumbert, 2000)

Traditionally, exemplar theory applies to speech perception, but various models have attempted to apply it to speech production as well. For example, a simplified model posits that when one is producing a sound, the entire memory category is activated and one exemplar is chosen at random. The probability that a particular exemplar is chosen is relative to that exemplar's resting activation level, or weight (Pierrehumbert, 2000). In contrast, the most widely cited and supported exemplar-based production theory says that the exemplars which get produced are not chosen at random from the full set; instead, all similar activated traces create a "generic echo," regressing towards the mean of the category. This echo is selected for production (Goldinger, 1997; Babel, 2009).

Either of these production models would allow for a change in one's perception and production of speech sounds, and could account for change in the regional-accent of a speaker over time. If a person relocates to a new dialect region and is presented with numerous exemplars representing the accent of that region, then the number of memory traces corresponding to that regional accent would increase, as would the weight of each of those traces. This would shift the mean of that category towards the regional variant, and over time could affect the talker's production. In this way, exemplar theory provides a very good base for explaining long-term accent change. Exemplar theory also inherently supports the idea that phonetic imitation continues throughout life, because it states that one's production of speech sounds is based on detailed memories of those sounds as produced by other people. A wonderful

element of exemplar theory is that it allows us to test the idea of phonetic imitation in a laboratory setting.

Because one's production of certain speech sounds, or tokens, is based on which exemplars for those tokens are most highly activated at the moment, exposure to words produced by a model talker will shift a participant's *immediate* production of those words. Their speech will shift from a generic version of the token towards a version that more closely resembles the speech of the model talker, because the exemplars that are most similar to those of the model talker are most highly activated in the moment of production (Babel, 2009). The term "immediate" is very important here, because with even a short delay, the speaker may revert to the generic echo (Goldinger, 1998).

Linguists have found that it is possible to trigger spontaneous phonetic imitation of single-word variables in socially-minimal situations using a shadowing method in which the participants listen to words spoken by the model talker and immediately repeat them. Generally, the purpose of these studies has been to answer various questions about episodic memory. Goldinger demonstrated that imitation increased with greater exposure to the exemplars in question and that imitation was related to the frequency of the words in question, meaning that common words (those with more stored exemplars) were more stable and were not imitated to the same degree as words with fewer exemplars (1998). Babel (2009) and Phillips and Clopper (2011) specifically studied imitation of American English vowels and found that the first and second formants of participant's vowels moved towards those of the model talker. These results support exemplar theory, as did previous studies; but the most interesting finding in the studies was the fact that the imitation effect was not equal for all vowel tokens. Specifically, both studies found that low-vowels /æ/ and /ɑ/ were significantly imitated, whereas higher vowels were not.

This suggests that exemplar echoes and phonetic imitation may account for certain aspects of intraspeaker phonetic variation, but that phonetic imitation may not occur automatically and may be affected by other biases and variables which have yet to be explored in research on spontaneous phonetic imitation.

One such variable that is noticeably absent from many studies of spontaneous phonetic imitation (see Goldinger, 1998) and which is discussed only on a preliminary level in Babel's (2009) experiment is the effect that the social value of certain exemplars has on the likelihood that they will be imitated. This is strong bias that could affect language variation and is absent from exemplar-based theories of production. However, in order to relate the exemplar-based model of change to the dialect studies discussed at the beginning of this paper, the social value/stigma of regional variations is extremely important. The study of social situation-based linguistic adjustment at the level of the individual is known as "speech accommodation."

Social factors in accent change: speech accommodation

The study of speech accommodation explores speech style variation based on social situation. The term "style" refers to intraspeaker variation; that is, it refers to variation within an individual's speech, not variation between speakers (Bell, 1984). Style is comprised of syntactic, lexical, prosodic, and pronunciation variables. If one's linguistic memory is comprised of stored exemplars, then style variation would refer to the use of different exemplars in different communicative situations. Various hypotheses have been proposed that attempt to explain the reasons behind stylistic shifts and changes (Hudson, 1999). While a prominent sociolinguistic theory maintained that variation in an individual's style was based on the amount of attention paid to speech (Labov, 1994), psychologist Howard Giles and colleagues took a different

approach to the phenomenon with their model of accommodation, known as Communication Accommodation Theory (CAT) (Giles, 1975; Trudgill, 1986). In this theory, it is hypothesized that variation in a person's speech is based on a desire to emphasize or minimize the social differences between the speaker and their interlocutor (Giles, 1975). Therefore, a central assumption of CAT is that people modify their communicative behavior in order to achieve specific intrapersonal goals such as social approval or more effective communication; a further use of speech accommodation might be to signal in-group membership or non-membership (Willemyns et al, 1997). Bell (1984) further explored potential causes for speech accommodation. His opinions supported the idea that stylistic variation is a result of audience design. He discussed at length possible motives for convergence and divergence of stylistic characteristics, and even went so far as to begin a discussion about referee design. Referee design is speech accommodation not towards the speech your audience or interlocutor, but instead towards a third party to whom you are referring (Bell, 1984). Taken together, these ideas represent a theory of stylistic change that is very much routed in social motivations and a sense of in-group or out-group membership. Speech accommodation is something which only happens in interactive settings where there is a communicative goal.

Giles' and Bell's theories of speech accommodation as a social psychological phenomenon have been supported by various studies. For example, Coupland (1980) found that in a Cardiff travel agency, an employee's accent routinely became more standardized (that is, they shifted from a local Cardiff accent towards an SSBE accent) when speaking with clients over the telephone and less so when speaking with coworkers in a more informal setting; this could represent nothing more than an attention to formality which triggered a change in speech style. However, in the course of his investigation Coupland also found instances where the

employee's style shifted towards the accent of her client, even when this accent was a non-standard accent. This occurred when the employee was attempting to smooth over a situation and, possibly, to create a sense of affinity. Thus, the research pointed towards the distinct possibility of interactional motivations for phonological style-shifting (Coupland, 1980). Another study which supported the interactional factor in style-shifting was carried out in Australia by Willemyns et al. (1997). They found that job applicants in a job interview would tend to shift towards broader Australian accents if the interviewer had a broad accent. The degree and likelihood of shift between participants was dependent on certain sociolinguistic variables, but again the researcher's general findings supported a theory of interaction and goal-based speech accommodation. A third example which could be used to support CAT is discussed by Hay et al. (1999), who found distinct evidence for referee design in the communicative accommodation of Oprah Winfrey. When the talk show host was introducing African American guests on her show, she tended towards vowel variants characteristic of African American English Vernacular as opposed to those of Standard American English, which is her typical dialect. The authors proposed that a possible reason for this shift could be a subconscious accent convergence to show affinity to her guests. In this case there is particular evidence to support this idea, because Oprah was shifting to a stigmatized-dialect (which she herself has publicly criticized) and was not actually interacting with the person whose speech she was accommodating towards (Hay et al, 1999).

These theories and studies have demonstrated that accents, or at least individual sound segments which are characteristic of accents, may change momentarily based on a desire to minimize differences between oneself and those with whom or about whom they are speaking; authors have attempted to apply this idea to long-term accent change as well. In their study of

Northern English students in London, Evans and Iverson (2007) hypothesized that changes in regional features were related to long-term social accommodation more than to simple exemplar-based changes in mental lexical representation; this reasoning fits with the fact that some students were found to have maintained their Northern English accents, possibly out of regional pride or as a method to signal group-membership. Furthermore, Evans and Iverson found that the NE speakers were more likely to shift their accents in sounds that were considered "stereotypical" of a Northern accent by both Northern and Southern speakers; however, they were less likely to accommodate sounds which they did not perceive as being stereotypical Northern features.

Both laboratory-based, spontaneous imitation and social-based accommodation have been shown to occur frequently and affect different elements of speech to different extents. However, a key difference between short term, situation based accommodation and the long term accommodation is the fact that people in the "long term" situation employ the new dialect features in various environments and with various interlocutors. This is not to say that these features become non-variable; however, the new pronunciations may cross situational boundaries, and that makes it more difficult to attribute them to social goals. The likeliest scenario is that both social accommodation and phonetic imitation occur naturally, and some interaction of the two influences long-term accent change.

The current study

Embarking most directly from the studies of Babel (2009) and Phillips and Clopper (2011), this study will be comprised of a lexical shadowing task focused on spontaneous phonetic imitation of American English vowels. The model talker will represent a non-standard,

regional American accent that is characteristic of Minnesota and Northern Wisconsin. The most notable feature is the raised and diphthongal quality of the low front vowel /æ/. This is the vowel found in words like “bat” and “back.” In this region, the vowel commonly takes the form of /æ͡ɛ/ (Allen, 1976), and it is particularly high before the voiced velar stop /g/. For many speakers, /æ/ that comes before /g/ (hereafter referred to as [æg]) has become merged with the vowel / e͡ɪ/. This means that the word “bag,” for some speakers of Upper Midwestern English, features a vowel which is homophonous to the vowel in “bake.” Because speakers of American English have the vowel / e͡ɪ/ in their native inventory, and because there is no phonotactic constraint against the production of / e͡ɪ / before /g/, it is plausible that people could pick up this pronunciation over time, and it is also plausible that we may see evidence of some degree of [æg] raising in a lexical shadowing task. However, because this raised vowel is also very marked as a non-standard variant, and because it intrudes into the / e͡ɪ / vowel space, it is possible that a social or regional bias or “stereotyping” would prevent speakers from adopting the pronunciation over time or in the shadowing task. Such biases would probably differ between subjects, and the differences could be based in part on where the participant is from and how much past experience the participant has with the accent. I expect that those participants who clearly identify the accent as distinct from the “norm” or who are previously unfamiliar with it will show less imitation than those who have more experience with the accent or who perhaps do not detect, consciously, the vowel raising. Generally speaking, the goals of this study will be to see whether 1) Speakers spontaneously imitate towards sound variations which do not exist in their variety of English, 2) whether length or degree of previous exposure correlates with the level of spontaneous imitation, and 3) whether social perceptions of the accent correlate with the level of spontaneous imitation in a non-interactive setting.

Method

Participants

Twenty-two undergraduate students from Macalester College and Ball State University participated in this study (13 women, 9 men; mean age=20.09 years, age range = 17-23 years). Participation in this study was voluntary. All participants were native speakers of American English and had no speech or hearing impairments; they were naïve as to the hypothesis under investigation and did not know the model talker. The participants represented three geographical dialect regions: six were from the Upper Midwest (the region including Minnesota, Wisconsin, and the Dakotas), seven were from the Lower Midwest (the region made up of Iowa, Nebraska, Kansas, Missouri, Illinois, Indiana, Ohio, and Pennsylvania) and seven were from the West Coast (Washington, Oregon, and California.) Two participants were excluded from analysis because their linguistic history made it difficult to classify them in a region. Regional dialect boundaries were based on Labov, Ash, & Boberg (1997) and Clopper (1994).

Materials/Apparatus

Participants were seated at a computer and were asked to wear Sennheiser HD 215 noise-blocking headphones. They were seated so that their mouth was roughly 8 inches from Dynex USB microphone. The experiment consisted of three types of stimuli. The first was written words presented on the computer monitor. The word list included a total of six words in which /æ/ preceded /g/ and three words in a control condition where /æ/ preceded /d/. Additionally, there were 21 filler words which varied in number of syllables and stressed vowels. A full list of words used can be found in the appendix.

The second stimulus was auditory; a one-minute story which prominently featured [æg] words, e.g. “bag” and “magazine.” A full transcript of the story appears in the appendix. The

story was read by a female model talker who spoke with a characteristically Minnesotan accent of the type described earlier. Crucially, her [æg] vowel proved to be significantly higher than her [æd] vowel (f1= 583.83 Hz vs. 732.67 Hz). The model talker was an 18-year-old undergraduate at Macalester College.

The purpose of the story task was to reduce comprehension difficulties and give context to the words that participants would be exposed to in task three, the shadowing task. It was a concern that words in isolation could simply be interpreted as mispronunciations instead of as typical pronunciations in a non-standard dialect; this could lead to brief confusion or comprehension difficulty and therefore difficulty in the shadowing task. Previous studies that examined accented-speech comprehension found that exposure to a short story read in the accent greatly reduced comprehension errors and processing costs (see Gass & Varonis, 1984). Another concern in this study was that a lack of context for the words in the shadowing task could lead to "bad maps" for those people who were unfamiliar with the accent. For example, if someone had never heard this accent before, then "flag" [fle͡ɪg] could easily be mapped onto "vague" [ve͡ɪg] or "flake" [fle͡ɪk] if it was uttered without phonological and semantic context, and this would lead to an incorrect shadow. Finally, hearing an accent in its entirety is useful for identifying it and/or forming socially-based value judgments about it, which we hoped to examine in this study. Therefore, by providing a prose passage, participants got a full picture of the speaker's accent and were able to hear that /æ/ raising was a consistent element of the model talker's phonology and not simply a mispronunciation.

The last task was the lexical shadowing task, which featured the same list of words as in the first task, spoken by the model talker from the second task. These words were presented in an auditory fashion.

After the experiment participants were asked to complete a questionnaire which asked for background biographical information from participants, as well as some questions about their perception of the accent of the model talker. The questionnaire can be found in the appendix.

Procedure

Upon arriving for the experiment, participants were asked to sit in front of the computer and center themselves over the microphone. The experiment was presented in a PowerPoint presentation; participants were instructed that it was automatically timed and that they should not press any buttons on the keyboard or mouse. All further instructions were presented on screen, so that conversation with the researcher would be limited. This was done in order to achieve the least biased results possible. Given the nature of the study, it was important to avoid the possibility of the researcher's own accent having an effect on the participant's baseline pronunciations or processing of the model talker's accent (the likelihood of such confounds is described in Hay, Drager, &Warren, 2010). Participants' responses were recorded using Audacity 1.2.6 software (distributed under GNU General Public License).

All on-screen instructions and visual stimuli were written in white, size 44 font on a navy blue background. Thirty seconds after the participants were notified that the experiment was beginning, instructions for the first task were displayed on screen. They read: "You will see a series of words presented on screen. When each new word is presented, please say it out loud. Each word will be displayed for 2 seconds. The task will begin in 10 seconds."

Once the task began, words were presented one at a time in random order. After reading all 30 words, there was a five second delay before instructions for the second task were presented. Participants were given 10 seconds to read the instructions "you will now listen to a story called "Maggie and Moe" before the story started.

A five second delay followed the story, and then the instructions for task 3 appeared: "You will hear a series of words. As soon as you hear each new word, please say that word out loud. Words will be presented every 2 seconds. The task will begin in 10 seconds." The words were presented one at a time in random order, as in Task 1. When this task was finished, participants were instructed to see the researcher for instructions in order to complete the questionnaire.

In order to analyze the recorded data, the first two formants of the every [æg] and [æd] token were analyzed using Praat Software (Boersma & Weenink, 2010). Formant measures were taken at the midpoint of the /æ/ in each word. Measurements were done by hand using the "formant listing" tool in Praat.

Results

Participants were considered to have shadowed correctly if the word that they uttered was the same as the one spoken by the model talker. If the participant said a different word than the one uttered by the model talker, this was counted as an incorrect response. An example of an incorrect response would be if the model talker said "flag" and the participant responded, "vague." The rate of accuracy for [æg] words in the shadowing task was 97.5%, with 117 out of 120 words being shadowed correctly. The rate of accuracy in the control condition was 96.7%, with 58 out of 60 words shadowed correctly. Incorrect responses were discarded in analysis. A further three [æg] tokens were discarded from analysis because it was unclear whether the participant was shadowing correctly. In each of these three cases, the participant had maintained a low level of imitation across [æg] words before raising their vowel all the way to the height of the model talker's on only one token. Though perhaps this is showing a true imitation effect, it more likely is showing that either a) the participant did not understand what word the model

talker was saying, and therefore just repeated the same sounds she made, or b) there was a lapse in attention that caused the participant to repeat the sounds without even considering what the word was. Because of this ambiguity these tokens were discarded.

The average F1 value of baseline [æg] productions across all participants was 848.36 Hz (min= 574.10, max= 1007.00, SD=100.14); the average F1 value of shadowed [æg] production was 769.23 Hz (min= 546.35, max= 939.67, SD=102.72). The average change in Hertz from the baseline condition to the shadowing condition was -79.13. A repeated measures t-test comparing participants' average baseline pronunciations to their average shadowed pronunciations found a significant effect for shadowing, with $p < .001$ ($t=7.23$, $df=19$). Figure 4 provides a graphical representation of these results, displaying a before-and-after of the participant's pronunciations in terms of both F1, which represents vowel height, and F2, which represents vowel backness.

The average F1 value of baseline [æd] productions was 880.68 Hz (min=685.22, max= 1066.80, SD=87.86); the average F1 value of the shadowed production was 845.29 Hz (min= 648.42, max= 1060.32, SD=103.06). The average change in Hertz from the baseline condition to the shadowing condition was -35.29 Hertz. A repeated measures t-test comparing the baseline pronunciations with shadowed pronunciations found a significant effect for shadowing, with $p =.048$ ($t=2.109$, $df=19$). A graph representing the changes between the baseline and shadowing conditions for the [æd] vowel can be seen in Figure 5.

There was not a significant difference between baseline pronunciations of [æd] and [æg] ($p=.161$, $t=1.460$, $df=19$). However, there was a significant difference between the shadowed pronunciations of [æd] and [æg] ($p=.008$, $t=2.945$, $df=19$). A phonetic distinction between these two vowels is consistent with the accent of the model talker.

An additional control analysis that was carried out concerned the F2 values in the baseline and shadowing conditions. The model talker's [æg] F2 was higher than all participants', meaning that her vowel was produced farther forward in the mouth. Given the results of the F1 analyses, it might be expected that a general trend of /æ/ fronting might occur in the shadowing task. However, statistical analyses showed that the average change in F2 from the baseline to shadowing conditions was not significant (p=.147).

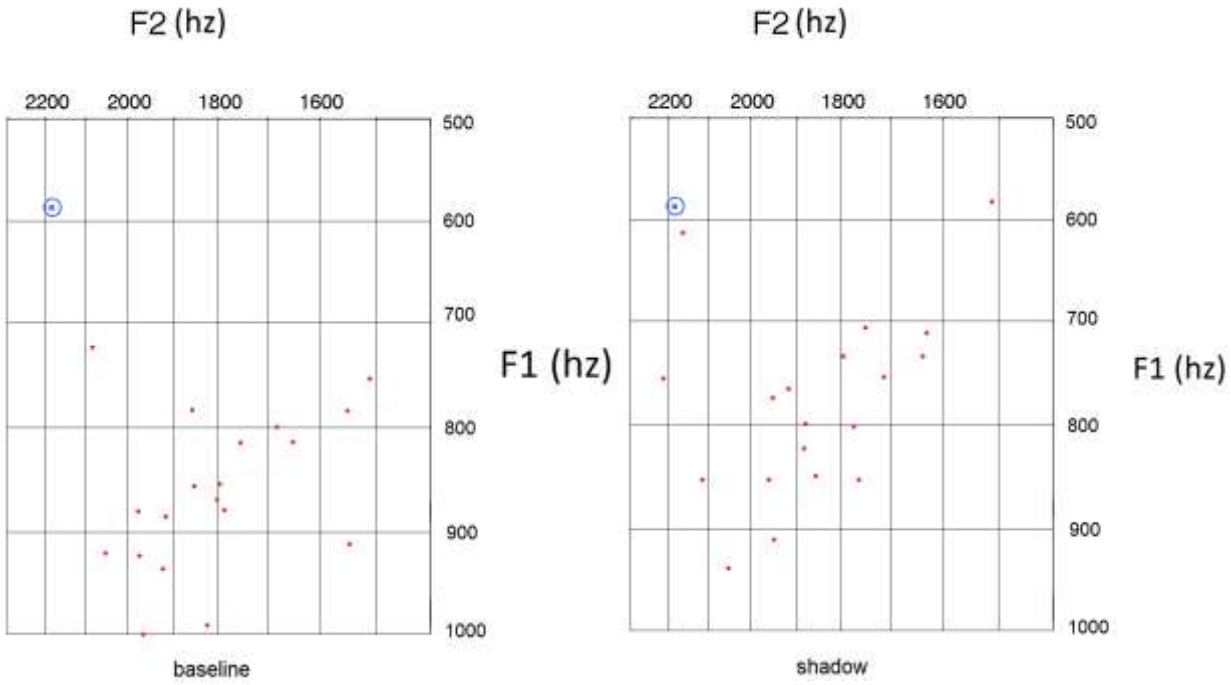


Figure 4. [æg] pronunciation in the baseline and shadowing conditions. The blue target represents the model talker's pronunciation. Each red dot corresponds to the pronunciation of a single participant. The general trend was for the participants' pronunciations shadowed pronunciations to be higher (have a lower F1) and are farther fronted (have a higher F2) than their baseline pronunciations.

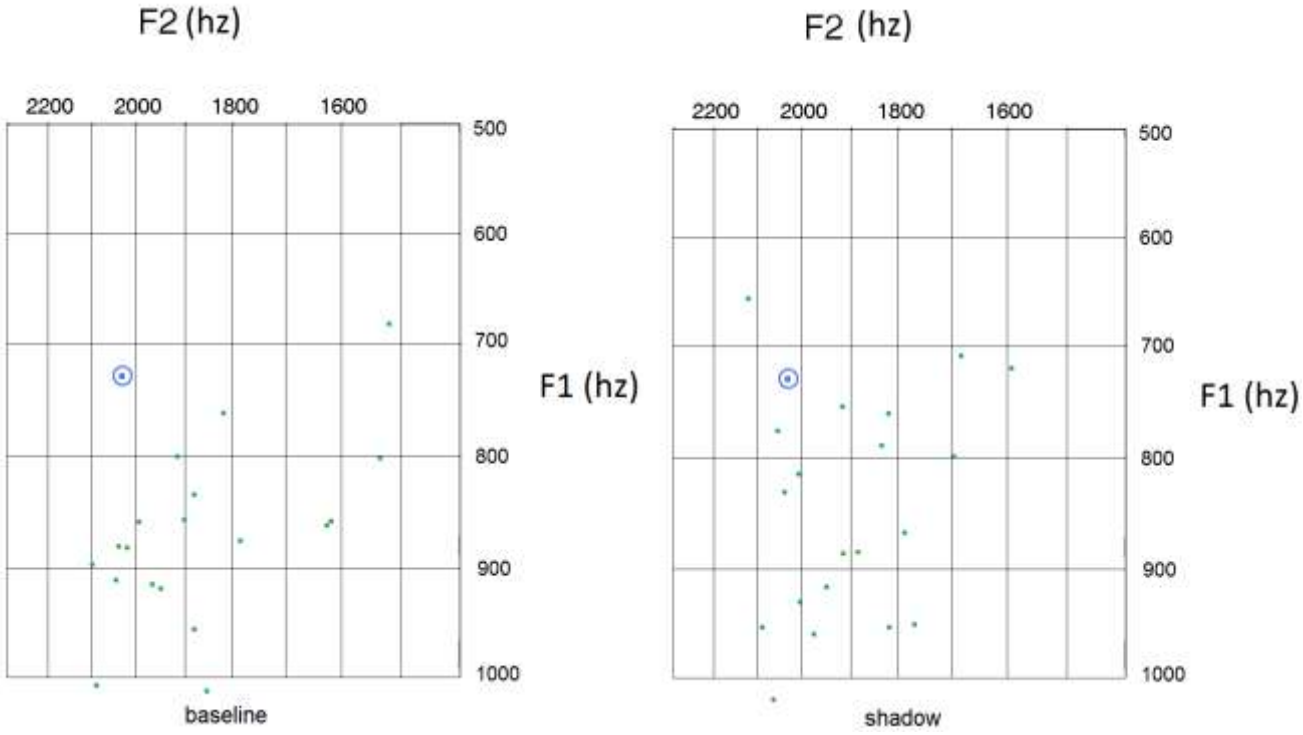


Figure 5. [æd] pronunciation in the baseline and shadowing conditions. The blue target represents the model talker's pronunciation. Each green dot corresponds to a single participant's pronunciation. There was again a significant effect for shadowing, though there was less uniformity in the "trajectory" of the vowel change between the baseline and shadowing conditions.

In order to examine these results in more detail, a series of one-way ANOVAs were carried out to compare various background and social features with participants' rate of imitation. The results showed that participant's self-reported familiarity and experience with this accent had no relationship with their level of imitation. Surprisingly, participants' home region and gender also did not correlate with the extent to which they imitated the vowel sounds.

In addition to collecting data about participants' regional backgrounds and self-reported experience with the accent of the model talker, the questionnaire also asked them to select from a

list of traits those that described the accent of the speaker. Participants were instructed to select as many as they wanted from a list of nine (see the questionnaire in the appendix for the complete list.) The words selected by the participants did not seem to relate in any way to their level of imitation; most importantly, there was no difference in the imitation rates between speakers who viewed the accent as "standard" vs. "non-standard."

Figure 6 shows the frequency that each descriptive term was chosen across all participants, though it should be noted that there were regional differences. For example, five out of seven of the participants from the Lower Midwest labeled the accent as "rural," while only two out of seven West Coasters did (and none of the people from the Upper Midwest did). Also, participants from the Lower Midwest were more likely than the other groups to label the accent as non-standard. 50% of the Upper Midwesterners rated the accent as standard and educated (see Figure 7). This could show a different perception of the accent by different groups, but more likely it shows that home region and life experience may affect that terms used to classify different accents. This type of variation needs to be taken into consideration when drawing conclusions about social biases and perceptions of dialect.

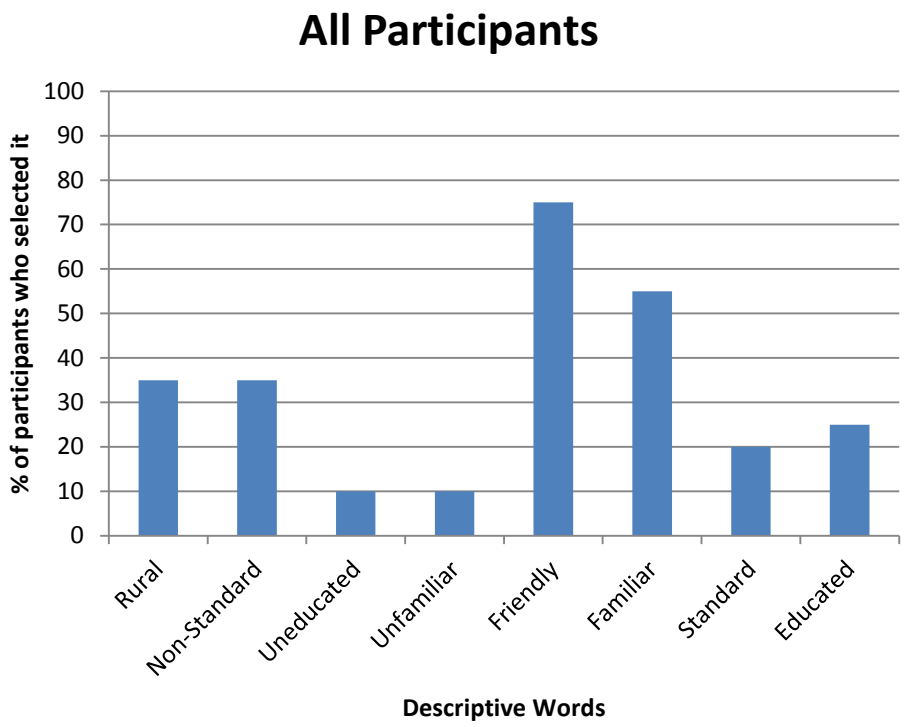


Figure 6. The descriptive words chosen by the participants to reflect the accent of the model talker, across all participants.

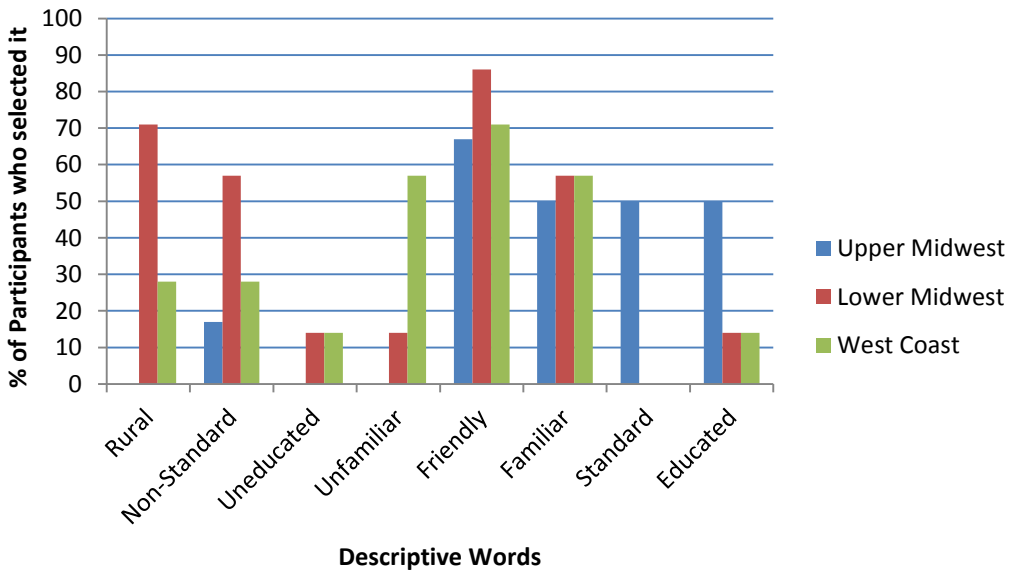


Figure 7. The descriptive words chosen by the participants to reflect the accent of the model talker, separated by participant home region.

Discussion

These results show strong evidence for spontaneous phonetic imitation. The data suggests that spontaneous phonetic imitation occurs even across regional boundaries. The imitation itself, in terms of average change in F1, was only slight. On average, participants raised their [æg] by only 80 Hertz despite the fact that the model talker's F1 was almost 300 Hertz higher than the average F1 of the participants. Generally, 80 Hz is an inaudible difference, especially to untrained ears. This shows that the imitation was probably subconscious and truly an effect of the brief exposure that participants had to the model talker's speech.

When asked after the experiment if they had any concerns or issues while carrying out the tasks, several participants (who were still naïve as to the goal of the study) noted that they felt themselves imitating and attempted to stop and produce the tokens in their "normal" way. Because shadowing tasks of spontaneous phonetic imitation are intended to create pronunciation changes that are completely subconscious, this type of comment highlighted both some positive effects and potential issues with the experiment. On the positive side, it showed that participants who were not attempting to mimic the speech of the model talker still imitated significantly-enough to notice it in their own speech. However, after they noticed the imitation they then made a conscious decision about how to pronounce the rest of the words. Obviously forcing participants to make a decision about how to speak was not the intent of the task, and the fact that some participants did make that decision could perceivably affect the results of the study. It is possible that additional participants also had this same experience but failed to mention it. Additionally, there was no way to control how people interpreted the instructions of the third task, "As soon as you hear each new word, please say that word out loud." Though the results do not suggest that anyone purposefully mimicked the accent of the model talker, it is possible that the instructions could have been interpreted to mean, "please say that word *as you would say it*."

For these reasons, it is unclear how many of the participants made a conscious decision to attempt to speak in their own accent, rather than mindlessly repeating the words that they heard from the model talker. However, the fact that there was significant imitation despite the possibility that some (or all) participants were making a conscious decision about how to speak is very encouraging, showing that the subconscious imitation effect caused by exemplar-activation may be strong enough that it persists despite focused intention to avoid it.

The most intriguing, and perhaps most meaningful, result of this study was the fact that a phonetic distinction between [æd] and [æg] did not exist in the participants' baseline speech, but was found in the shadowing condition. This shows that despite the fact that actual imitation was not extreme, the participants were imitating the two vowel variants differently and beginning to show patterns of pronunciation that more closely resembled the model talker's.

A surprising finding is that the average amount of imitation on the regionally-marked vowel [æg] exceeded that of control vowel [æd]. The reasoning for this may be that participants had "farther to go" when shadowing these sounds. That is, the [æg] vowel was significantly higher than the [æd] vowel, and therefore participants were physically able to alter their pronunciation to a greater degree. Furthermore, the [æg] of the model talker was so far removed from everyone else's that the direction with which they needed to adjust was clear; the vowel was higher and farther forward. With [æd], the situation was a little more complicated. In the baseline condition, one participant actually had a higher pronunciation, and several participants had pronunciations that were equally far forward. For that reason, formant values at the vowel nucleus are perhaps not the best place to look for significant and meaningful imitation. An element of the vowel which was not analyzed fully in this study was its diphthongal quality. In the model talker's speech, the vowel in [æd] is pronounced as the diphthong /æ͡ə/, which is

common in Minnesota (Allen, 1976). Comparing the trajectory of the formants throughout the duration of the vowel sound would allow one to note a decrease or increase in diphthongization between the baseline and shadowing pronunciations of the participants. If subjects for whom /æ/ is normally a monophthong began producing it as a diphthong in the shadowing task, this would show another type of subtle imitation. This type of analysis could also be carried out for the [æg] condition. Further studies of cross-dialect phonetic imitation should focus on a variety of sounds and sound features that are regionally marked in different ways; for example, one might consider suprasegmental factors such as length, duration, and pitch when studying vowels. By using a variety of measures results such as the ones described here would be more robust and convincing.

The fact that significant imitation was found on F1 and not F2 also poses an interesting question. When looking at the plotted data it does seem as though the general trend was to imitate F2, even if this effect was not significant. Returning to an exemplar-based explanation for spontaneous phonetic imitation, it is possible that the model talker's F2 was not as distinct as her F1 in terms of distance from the "generic echo" of the exemplar cloud of most participants. Therefore, it simply did not affect their pronunciation as much as the raising did.

The hypotheses set forth at the beginning of this experiment were not all supported. It is particularly interesting that social factors such as gender, home region, experience with the accent and time spent in Minnesota did not seem to be associated with the rate of imitation. In previous studies, these social factors have played a part in both spontaneous imitation and social-based accommodation (Willemyns et al, 1997; Babel, 2009). It is unclear why we did not find similar results.

It is important to note that over the course of the experiment there were some unusual findings that drew into question the perceptual nature of the Minnesotan /æ/ raising. Not only are these findings interesting, as the Minnesotan dialect has received relatively little research in the past, but they are extremely important as they could impact the conclusions drawn from this experiment. One of the most interesting things that occurred during the study was three instances of hyper-correction. One of the filler words in the study was "bagel," pronounced by the model talker in the standard way, as [beɪgəl]. All participants also pronounced it this way in the baseline condition. However, in the shadowing task three separate participants pronounced it as [bægəl] before correcting themselves and saying the word in the standard way. In this situation one has to wonder whether or not the subjects had subconsciously created a rule saying that when they heard [eɪ] before /g/ (as in the model talker's pronunciation of [æg]), then it was to be corrected to [æ]. If this is indeed the case, then it is safe to say that they interpreted the Minnesota accent as featuring a phonological replacement—/eɪ/ replaces /æ/ before /g/--instead of interpreting the difference as merely a raising of the vowel /æ/. This study has been assuming that the standard [æd] sound and the raised [æg] sound are merely allophones of /æ/, as this was the viewpoint of the model talker about her own speech. However, it seems that not all participants recognized it as such. It would be interesting to study the nature of this vowel more closely. Providing more instances in which hyper correction was possible could have helped illuminate the issue more clearly in this study. For example, including words like "vague" and "plague," in which /g/ follows /eɪ / in standard pronunciation, would have allowed us to test the robustness of the hypercorrection effect.

The reason why the perceptual nature of this vowel is important is because it could be related to the likelihood of imitation. Asking someone to listen to the vowel /eɪ/ and then to

repeat back the vowel /æ/ could potentially have different effects than asking someone to listen to a particular exemplar of /æ/ and then repeat back the same phoneme. In the former case, participants would be mapping the incoming sound onto one exemplar cloud and then creating their response from another. Therefore, even if imitation did occur it might not be explained in the same way. An interesting correlation to look for might not be participants' experience with the accent in question, but their phonological perception of the sounds of the accent and the imitation that occurs in their speech.

Another topic for future studies to explore is the time-course of the imitation effect. That is, how long do participants maintain their modified pronunciation after the initial exposure? Exemplar theory may suggest that the more exposure a participant has had to an accent, the longer the imitation effect will persist, because they have more, heavily weighted memory traces related to that accent. Studying the time-course of cross-dialect imitation would help to support exemplar theory as an underlying factor in regional accent change.

Lastly, another analysis that might be carried out is looking at imitation of specific words. It would be interesting to know if words that appeared earlier in the shadowing task were imitated to a greater degree, because many participants mentioned that they noticed themselves imitating and tried to correct it as the shadowing task progressed. Also, it would be interesting to see if words that appeared in the story *and* the shadowing task got imitated to a higher degree than words that only appeared in the shadowing task. Because the participants had more recent exemplars of the words that appeared in both the story and the shadowing task, one could imagine that they might be especially imitated. However, there would only be three words in each group in this analysis, and therefore no real conclusions could be drawn.

The overall results of this study support previous research which investigated spontaneous phonetic imitation in non-interactive tasks. By using a lexical shadowing task in which there were no social goals or interaction between the participant and the model talker, we were able to control for many of the possible motivations underlying speech accommodation theory. However, we were able to reintroduce a social element by using a non-standard, regional dialect. The study found that regionally marked variants were significantly imitated, regardless of the linguistic background of the participants and their attitudes towards the accent of the speaker. This result suggests that imitation may be a fairly automatic process, at least for the vowel /æ/. This study also raised questions about the perceptual nature of Minnesotan /æ/ raising, a topic which merits further study.

In conclusion, the process of spontaneous phonetic imitation is one which could very well underlie long-term accent change in individuals. This study has shown that spontaneous phonetic imitation is not restricted to within-dialect variables, and that even brief exposure to a regional accent is enough to slightly modify participant's pronunciations. Though this experiment only focused on one regionally-salient vowel variable from one specific accent, the results support further research on this topic. Namely, this type of experiment should be extended onto studies of other sounds and other dialects. This would allow us to look comparatively between different types of regionally-salient variables and different accents, so that we could draw more conclusions about the extent to which spontaneous phonetic imitation occurs across dialects.

References

Allen, H.B. (1976). *Linguistic Atlas of the Upper Midwest* (Vol. 1). Minneapolis, MN: University of Minnesota Press.

Babel, M. (2009). Selective vowel imitation in spontaneous phonetic accommodation. *UC Berkeley Phonology Lab Annual Report 2009*, 163-194.

Bell, A. (1984). Language style as audience design. *Language in Society, 13*(2), 145-204.

Boersma, P. & Weenink, D. (2010). Praat: doing phonetics by computer [Computer program]. Version 5.3.13, retrieved September 2010 from http://www.praat.org/

Clopper, C.G (2004). *Linguistic Experience and the Perceptual Classification of Dialect Variation* (doctoral dissertation). Retrieved from http://www.ling.ohiostate.edu /~cclopper/ clopper_ thesis.pdf

Coupland, N. (1980). Style-shifting in a Cardiff work-setting. *Language and Society, 9*(1), 1-12.

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

Gass, S. & Varonis, E.M. (1984). The effect of familiarity on the comprehensibility of non-native speech. *Language Learning, 34*(1), 65-89.

Giles, H. & Powesland, P. (1975). Accommodation Theory. In N. Coupland & A. Jaworski (Eds.), *Sociolinguistics: A reader and coursebook* (1997), 232-239.

Goldinger, S.D. (1997). Words and voices: perception and production in an episodic lexicon. In K. Johnson & J.W. Mullennix (Eds.), *Talker Variability in Speech Processing* (33-66).

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

Hay, J., Drager, K. & Warren, P. (2010). Short-term exposure to one dialect affects processing of

another. *Language and Speech*, 53(4), 447-471.

Hay, J., Jannedy, S. & Mendoza-Denton, N. (1999). Oprah and /ay/: Lexical frequency, referee design and style. Paper presented at the 14th International Congress of Phonetic Sciences, San Francisco, CA. Paper retrieved from http://www.ling.ohiostate.edu/~jannedy/DOCS /icphs.html

Labov, W. (1994). *Principles of Linguistic Change: Internal Factors*. Cambridge, MA: Blackwell Publishers.

Labov, W., Ash, S. & Boberg, C. (1997). *A National Map of Regional Dialects of American English.* http://www.ling.upenn.edu/phono_atlas/NationalMap/ NationalMap.html

Ladefoged, P. (1975). *A Course in Phonetics*. San Diego: Harcourt Brace Jovanovich Publishers.

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.

Phillips, S. & Clopper, C.G. (2011, May 23). Implicit imitation of regional dialects: acoustic analysis. Paper presented at the 161st Acoustical Society of America Meeting, Seattle, WA. Abstract retrieved from http://www.aip.org/ asalaypapers2011 /Phillips.html

Pierrehumbert, J.B. (2000). Exemplar dynamics: word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (137-159).

Trudgill, P. (1986) *Dialects in Contact*. Oxford: Basil Blackwell Ltd.

Willemyns, M., Gallois, C., Callan, V. & Pittam, J. (1997) Accent accommodation in the job interview: impact of interviewer accent and gender. *Journal of Language and Social Psychology*, *16*(1), 3-22.

Appendix

[æg] (experimental)	[æd] (control)	Other vowels (filler)
Magazine	Mad	Bagel
Wagon	Badger	Maid
Bagpipes	Dad	Moon
Dagger		Mow
Flag		Hippo
Zigzag		Phone
		Pencil
		Flame
		Soup
		Bone
		Ice cream
		Book
		Bead
		Cow
		Goal
		Watch
		Tuba
		Boot
		College
		Dog
		Puppet
		Snow

Appendix 1. Wordlist used for shadowing.

Twins Maggie and Moe were an interesting pair. Moe spoke six languages. Maggie played the bagpipes and boasted the world's largest collection of flags. After growing up in tropical paradise, they moved north for college. They started their college careers with many brag worthy accomplishments, but they hit a snag when it turned out that they had forgotten to prepare for one thing: winter weather. When the air froze and the first big snow fell, Moe handled it like a pro but his twin panicked…Moe literally had to drag her outside so she could go to class. As the winter went on, Maggie spent more and more of her time indoors practicing her bagpipes and organizing her flag collection, while Moe zigzagged across campus on skis. He soon found himself nagging his sister to leave the dorm more often, claiming she was falling behind in class work. However, she simply gagged at the thought of traipsing through 2 feet of snow.

Appendix 2. The story participants heard during task 2.

Appendix 3. Survey given to all participants in order to gather background data.

Biographical Survey

1) What is your age?_____

2) What is your gender? _____

3) Are you a native speaker of American English? _____

4) Do you have any speech or hearing impairments?_____

5) Where did grow up? (you may list multiple places; please indicate in which places you spent the most time)

6) When you were growing up, were your primary guardians native speakers of American English? (Please answer for each guardian.)

7) Where did your guardians grow up?

Post-Listening Survey

Please answer the following questions about the accent of **the speaker in the recording**. For the purposes of this survey, a standard/neutral accent is one which is not identifiable as belonging to a particular geographic region.

1) Given the regional choices (A-G) on the map, where would you place this speaker's accent? _____

2) How confident are you in this answer?

1	**2**	**3**	**4**	**5**
(not at all)				(Completely confident)

3) How familiar are you with accents from this region?

1	**2**	**3**	**4**	**5**
(not at all)				(Very familiar)

4) How similar does the speaker's accent sound to your own accent?

1	**2**	**3**	**4**	**5**
(not at all)				(very similar)

5) Please circle as many words as you feel describe this person's accent:

uneducated	standard	rural
sophisticated	familiar	non-standard
friendly	unfamiliar	educated

Please answer the following questions about your own accent.

1) In general, my American accent is quite standard/neutral.

1 (strongly disagree)	**2**	**3**	**4**	**5** (strongly agree)

2) My accent has become more neutral since coming to Macalester**.**

1 (strongly disagree)	**2**	**3**	**4**	**5** (strongly agree)

3) My accent has become less neutral since coming to Macalester.

1 (strongly disagree)	**2**	**3**	**4**	**5** (strongly agree)