RICE UNIVERSITY

Imitation, Awareness, and Folk Linguistic Artifacts

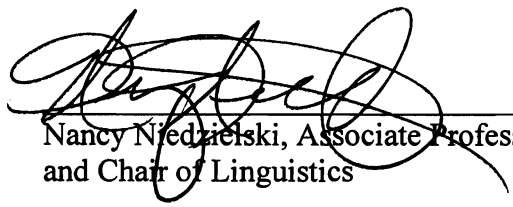
by

Elizabeth Gentry Brunner

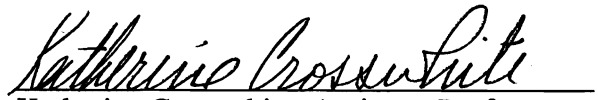
A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

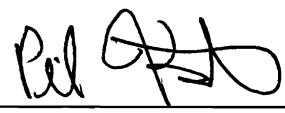
APPROVED, THESIS COMMITTEE:



Nancy Niedzielski, Associate Professor
and Chair of Linguistics



Katherine Crosswhite, Assistant Professor
of Linguistics



Philip Kortum, Professor-in-the-practice
and Faculty Fellow of Psychology

HOUSTON, TEXAS
SEPTEMBER 2010

Abstract

Imitation, Awareness, and Folk Linguistic Artifacts

by

Elizabeth Gentry Brunner

Imitations are sophisticated performances displaying regular patterns. The study of imitation allows linguists to understand speakers' perceptions of sociolinguistic variation. In this dissertation, I analyze imitations of non-native accents in order to answer two questions: what can imitation reveal about perception, and how are *folk linguistic artifacts* (Preston 1996) involved in imitation? These questions are approached from the framework of *folk linguistic awareness* (Preston 1996). By redefining the concept of salience according to the modes of folk linguistic awareness, I am able to more precisely consider how imitation reflects salience. I address both of these questions by eliciting imitations from speakers in which folk artifacts are present.

For my investigation, twenty speakers read a short passage in English. Ten were non-native speakers of American English (NNAE) and ten were native speakers of American English (AE). The AE speakers were recorded reading the passage in their regular voice and with two types of imitated accents: *free imitations*, which were spontaneously produced, and *modeled imitations*, which were produced

directly after hearing the NNAE speakers. Free imitations revealed folk linguistic artifacts, while modeled imitations were more reflective of the immediate target. Participants listened to the authentic and imitated accents and were asked to determine the accent and authenticity of each speaker.

I found that there was not a significant difference in the pitch and vowels between free and modeled AE imitations, which indicated that these aspects of imitations are largely based on folk linguistic artifacts. Listeners were able to determine which voices were authentic and which were imitated. Listeners were also able to identify the speakers' accents, perhaps aided by the folk artifact status of these particular accents. Listeners were better at identifying the accents of free imitations than modeled imitations, which suggested that listeners prefer imitations that are solely based on folk artifacts.

Overall, I found that imitation is a valuable tool for the analysis of speech perception. The modes of folk linguistic awareness are useful in interpreting imitations and understanding salience. This research shows that folk linguistic artifacts are the foundation of imitations and an important tool in perceptual categorization.

Acknowledgments

I'm grateful to many people for their help along the way. Thanks to my adviser Nancy Niedzielski, who read draft after draft and didn't complain about my constant phone calls. Thanks also to my other two committee members, Katherine Crosswhite and Phil Kortum, whose doors were always open.

A big thank you goes to all of the participants, and particularly the ones who took the survey for free. Thanks to Caroline Nganga for running the statistics that I couldn't.

I owe so much to Michelle Morrison and Andrew Pantos, who have been with me through every step of graduate school. They made the good better and the not-so-good bearable, and I treasure their friendship. Thanks also to my officemate Chris Koops, who was always up for discussing data. The ladies of the Rice Graduate Christian Fellowship helped me remember that there are things more important than a dissertation, and the importance of living intentionally, and I'm so grateful to have had their support and prayers throughout my time at Rice.

I have to thank my entire family for their support and prayers also. My parents in particular have encouraged me every step of the way, and my brothers kept their loving "You're *still* in school?!" teasing to a minimum.

A huge thank you goes to my amazing husband, Gregory. He was always patient and realistic about this process, and read through many drafts himself. He knew how to motivate me every step of the way. This is a better dissertation because of him, and more importantly, I am a better person because of him.

Finally, I am overwhelmed with gratitude to God for giving me this opportunity. I hope that in some little way my experience in graduate school has brought Him glory.

To my parents

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Imitations tend to be disregarded by sociolinguists as nothing more than inaccurate stereotypes. However, imitations may be sophisticated performances displaying regular patterning that can teach us much about listener perceptions and awareness. This dissertation investigates imitations of non-native accents in order to explore two questions. The first question is: What can imitation reveal about perception? Several researchers have suggested that imitation displays the features that speakers perceive to be salient (e.g. Trudgill 1986, Schilling-Estes 1995), yet this is a fairly simplistic outlook that does not fully reflect the various types of awareness speakers have. By focusing on perception through a lens of *folk linguistic awareness* (Preston 1996), we can interpret imitations (and listeners' evaluations of them) without overstating or understating their significance. We can also reframe the concept of salience in light of this awareness, and therefore distinguish a speaker's ability to modify a feature (control) from the salience of a feature (availability). The second question is: How are *folk linguistic artifacts* (Preston 1996) involved in perception? It appears that imitations are based, at least partly, on folk artifacts. Folk artifacts may also more broadly assist in categorization. I am able to address both of these questions by eliciting imitations in which folk artifacts are present.

In this research I seek to address these questions by looking at perception, and specifically awareness, from two different angles. The first is speakers' perceptions as revealed in their productions of non-native imitations; the second is listeners' perceptions as revealed in their evaluations of these imitations. Studying speaker perception will give us insight into whether free (spontaneous) imitations or modeled imitations (those immediately following hearing an authentic accent) are more similar to the target accent. Studying listener perceptions will allow us to determine whether listeners can accurately identify accents, whether they can distinguish between an authentic and an imitated accent, and whether they react differently to free and modeled imitations.

This chapter provides the background for the theoretical concepts of folk linguistic awareness, folk linguistic artifacts, and imitations. First, I present a history of how awareness has been approached by sociolinguists, concluding that the most complete framework is provided by Preston 1996's folk linguistic awareness model. Next, I introduce folk linguistic artifacts, which are conventionalized depictions of other language varieties. Finally, I present previous research that has been done on imitations, finishing with the connection between imitation and salience and the difficulties involved in studying imitations. I conclude the chapter with an overview of the rest of this dissertation.

1.1 Linguistic Awareness

Preston 1996 defines linguistic *awareness* as "the 'degree' of consciousness non-linguists have in general about language" (72). As such, awareness is

presumably a crucial part of language perception. Though perception has been somewhat neglected by sociolinguists in the past (Thomas 2002), awareness has been an oft-cited factor in sociolinguistic research. Awareness is in fact assumed to motivate particular sociolinguistic phenomena, explicitly in "social awareness" and implicitly in the form of attention paid to speech. However, this is an incomplete approach to awareness. It is worthwhile to review the role that awareness has played in the field of sociolinguistics, in order to fully appreciate the importance of Preston's contribution of a more nuanced understanding of folk linguistic awareness.

1.1.1 Sociolinguistic Approaches to Awareness

Sociolinguists have followed Labov (1966) in identifying two types of linguistic change based on speaker awareness: change from above the level of consciousness and change from below the level of consciousness. In change from above, speakers are aware of the variation; in change from below, speakers are initially unaware of the variation. Labov maintained that change from below was reflected by the social awareness surrounding a feature undergoing change (Labov 2001). This understanding of awareness refers to how cognizant a group of people is about a linguistic feature: how available that feature is to them. Labov proposed that a feature could reveal this social awareness in three ways: as an indicator, a marker, or a stereotype. A feature starts as an indicator, the variation of which no one is aware. As the feature continues to change, however, speakers start to become

Related to this was Labov's Vernacular Principle, which is perhaps one of his most influential contributions to the field of sociolinguistics: "the style which is most regular in its structure and in its relation to the evolution of the language is the vernacular, in which the minimum attention is paid to speech" (112). This primacy of unmonitored speech reflects Labov's assertion that "the most systematic and regular form of language is that of basic vernacular" (Labov et al. 1968:167). According to the Vernacular Principle, less attention to speech is associated with structural regularity and significance for linguistic change. The implication is that styles in which a great deal of attention is paid to speech are irregular and unreflective of linguistic processes. The focus on vernacular speech emphasizes little conscious monitoring, which we may interpret as a type of low awareness of speech.

Labov's emphasis on the vernacular led to a tacit understanding by most sociolinguists of "unconscious" speech as authentic (Eckert 2003, Bucholtz 2003, Coupland 2003). Bucholtz 2003 calls this "the ideology of linguistic mundaneness": the assumption that "the most authentic language is language that, from its user's point of view, is unremarkable, commonplace, everyday" (405). Coupland 2007 confirms "it's clear that variationist sociolinguistics has taken an ideological stance in favour of vernaculars, and that it has assumed that vernaculars are authentic speech products" (181). Along with this implicit understanding of vernacular speech as authentic comes the presumption that only authentic speech is worthy of investigation. Because of this emphasis on vernacular speech as authentic speech, particularly by American variationists, there are fewer sociolinguists who have

studied the supposedly less authentic monitored speech styles, such as performance speech. But there is a great deal that linguists can learn from looking beyond the vernacular to these more self-conscious styles.

In performance speech a speaker intentionally presents their interpretation of a certain language variety to an audience. Schilling-Estes 1998 defines performance speech as "that register associated with speakers' attempting to display for others a certain language or language variety, whether their own or that of another speech community" (53). She goes on to observe that "speakers highlight features of which they are most aware (whether at the conscious or unconscious level) when they give a speech performance" (77). For many years it was assumed that performance speech was not modified in a systematic way, and only vernacular speech showed structural regularities (e.g. Labov 1972); indeed, performance speech can be irregular at times. Performances may be more or less successful, and individual speaker skill at performing may vary. However, not all performances are irregular. Preston 1996 observes that a speaker who is more involved in their performance "dramatically enhances the effect (if not the accuracy)" (64). Further, studies such as Schilling-Estes 1998 and Evans 2002 have revealed that performance speech can exhibit regular patterning. Evans 2002 demonstrated that a non-Southern speaker could perform a Southern dialect, conforming his imitated vowels to the patterns of the Southern Shift. In a case study of Ocracoke English, Schilling-Estes 1998 observed that her informant's "performed /ay/ vowels display the same sensitivity to phonological conditioning as these vowels in non-performance speech, and so do not display the increased irregularity that Labov

maintains should accompany increased attention to speech" (69). Chun 2006 also concluded that "performed speech shares many of the patterns of 'unperformed' speech" (48). Performance speech can display structural regularities despite the speaker's increased monitoring.

There is a great deal of awareness evident in performance speech which cannot be sufficiently described using the attention to speech model or the social awareness model. Those approaches focus on the speech itself (what speakers do), but awareness is broader than that: it also encompasses what speakers know. As Preston 1996 puts it, "it is important not only to know what language is and how people use it but also what they think about it" (72). To this I would also add *how* people think about it; that is, how people perceive it. In order to understand these aspects of speaker awareness, we must step back and approach it from a different angle: that of the "folk."

1.1.2 Folk Linguistic Awareness

Folk linguistics is the study of language from the layperson's point of view, or "folk" view. This field seeks to understand language through reflections of the folk. It is related to the field of perceptual dialectology, which "has the goal of uncovering the folk's own understanding of different varieties, looking not only at evaluations of language varieties but also at how these varieties are categorized" (Lindemann 2005:189). Dennis Preston has been at the forefront of folk linguistic research (e.g. Preston 1989, 1996, 1999, Long & Preston 1999, Niedzielski & Preston 2000), and

aware of it, and their productions of the feature vary by social group and style; this signals that the feature has become a marker. If speakers become so aware of a feature that they can overtly comment on it, it has become a stereotype.

Labov's social awareness is a good start to understanding awareness, but it is ultimately insufficient because it does not fully address all that speakers are aware of in language. The social awareness of features is an aspect of the availability of features, and a valuable approach to understanding availability; however, as we shall see in section 1.1.2, while availability is an important facet of linguistic awareness, it is only one among several factors contributing to awareness.

Another sociolinguistic concept that is implicitly rooted in awareness is attention to speech.[1] This theory was originally proposed by Labov (1966, 1972), whose Principle of Attention stated, "Styles can be ordered along a single dimension, measured by the amount of attention paid to speech" (1972:112). Style was essentially defined by the amount of conscious monitoring of speech. The premise was that the less speakers paid attention to their speech production, the more likely they would be to produce speech that was more casual, natural, and regular. We might say that the goal, therefore, was for speakers to be "unaware" of their speech.

[1] The attention to speech concept has proven to be inadequate for two reasons, unrelated to awareness. The first is that it is difficult to operationalize attention to speech. While Labov 1972 took great care in creating a stylistic continuum based on situational formality, the attention to speech component was not quantified. As Rickford & McNair-Knox 1994 explain, "researchers usually found this method of distinguishing casual speech [based on attention] difficult to apply in an objective and reliable way" (238). The second problem with the attention to speech model is that it is not theoretically compatible with what sociolinguists have learned about style (Bell 1984:147-150; Milroy 1987:172-183; Rickford & McNair-Knox 1994). Stylistic variation springs from a great many factors besides formality, such as audience and identity; it cannot be described only in terms of attention paid to speech.

he maintains that linguists can learn about language by exploring the folk's language awareness. To do this, Preston 1996 proposes four *modes of folk linguistic awareness*. These four modes are availability, accuracy, detail, and control. *Availability* is how much or how little non-linguists talk about a variety. An aspect of language may be completely unavailable to a non-linguist; it may be available only after careful description and explanation by a linguist; it may be suggestible, which is to say that non-linguists will comment on a feature, though they will not bring it up; or it may be common and frequently discussed by the folk. *Accuracy* is whether what the folk say is correct. Crucially, even when the observation is incorrect, it may still provide valuable information. *Detail* is how exact non-linguists can get in describing features of a particular variety; this ranges from global to specific. *Control* is whether the non-linguists can perform a variety, or aspects of it.

Each of these four modes is made up of a continuum and is relatively independent of the others. For example, while speakers may frequently comment on a person's accent (high availability), they may be unable to comment on specific linguistic features (only global detail instead of specific). Similarly, a speaker may be able to accurately imitate an accent (high control) but be unable to explain what they modified within their speech (again, global detail). A layperson may notice a particular variety (high availability), but believe incorrect facts regarding it (low accuracy). Figure 1 presents a hypothetical chart of these modes, showing their independence. This figure uses the above-mentioned continua, and depicts the

following fictional example: A Spanish accent[2] is very available to the folk: it is the topic of overt comment. However, the folk have very little detail about this accent. Nonetheless, their comments have some degree of accuracy and they may be able to partially imitate it in conversation.

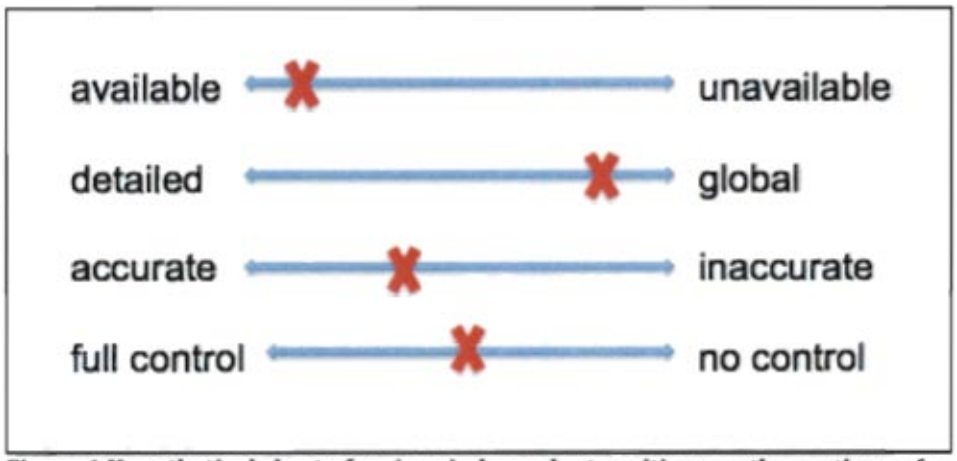


Figure 1 Hypothetical chart of various independent positions on the continua of modes of folk linguistic awareness. Adapted from Preston 1996:41.

This understanding of awareness is clearly different from the sociolinguistic approaches discussed in the previous section. While both are based on a type of "consciousness," Preston's folk linguistic awareness considers consciousness a type of knowledge, while Labov's attention to speech model considers consciousness a type of monitoring. Labov's social awareness is just one aspect (namely, availability) of Preston's more detailed understanding of awareness.

---

[2] Of course, there is no such thing as a single e.g. "Spanish accent." The English accent of individual non-native speakers of the same native language may differ widely, based on such factors as dialect (e.g., Argentinean versus Castilian, and the variations within each of those), education (e.g., was English learned academically or naturally? what variety of English was taught?), level of proficiency, experience, etc. Nonetheless, listeners tend to classify them as one general folk artifact and perceive them as such (Podbresky et al. 1990). Therefore, non-native accents will be treated as singular entities in this paper. This is similar to a linguist discussing Southern US English; while this accent varies by region and speaker, there are enough similarities that it can be considered an acceptable overarching category.

Preston's folk linguistic awareness research originally focused on regional dialects; the present work extends his concepts to non-native accents. It explores native American English (AE) speakers' imitations and perceptions of non-native American English (NNAE) speech. No other studies have investigated NNAE speech and imitations from an awareness stance, even though the folk linguistic awareness framework allows for a more nuanced interpretation of perception. Additionally, a folk linguistic awareness approach facilitates a clearer understanding of imitations of these accents. A speaker's ability to perform an imitation is reflected in the control mode. This control may be more or less accurate, and more or less detailed. This control may be independent of how available the target of the imitation is. The interaction of the control mode with the other three modes of linguistic awareness allows us to more precisely analyze imitations. Through folk linguistic awareness (hereafter often referred to simply as "awareness"), we are able to reflect on not only what speakers do, but also what they know.

1.2 Folk Linguistic Artifacts

Preston notes several factors that influence linguistic awareness, including formal training, correctness, publicity, and folk artifacts.[3] He contends that *folk linguistic artifacts* contribute substantially to a culture's linguistic awareness. Folk artifacts are the traditional shared impressions that a society holds about different language varieties. Markham 1997's explanation of caricatures as "conventionalised

[3] Williams et al. 1999 suggest affective factors be added to this list.

social representations" (82) of a group aptly describes folk artifacts. Often the features of folk artifacts have been accumulated more from previous generations and the media than from actual interaction with speakers of that variety. According to Preston, a folk artifact involves "a clear (and unmistakable) reference to (or imitation of) [a particular variety of] English, one which carries with it the immediate speech community recognition of exactly what it is the performer is about, including a clear understanding of what symbolic characterisation is intended" (1996:60).

We may say that folk artifacts are similar to stereotypes, although it is important to differentiate the two. Speakers may vary in their linguistic awareness of a folk artifact, no matter how strongly it may be stereotyped. Thus in this case we may say that *stereotype* is limited to referencing availability, which is only one of the modes of folk linguistic awareness. Stereotypes may also reference accuracy, with a connotation of inaccuracy, while folk artifacts may be anywhere on the accuracy continuum. Since my goal is to understand perception from a folk linguistic awareness standpoint, this dissertation intentionally focuses on folk artifacts, as opposed to stereotypes.

Societies can hold a variety of folk artifacts. These may be based on a number of social features, such as age, race/ethnicity, socioeconomic class, profession, and geography. Geographic region is a prime basis for artifacts; we can see this in strong American artifacts of Southern speech and New York City speech, for example. This is also the case with non-native accents: Americans hold a variety of artifacts for non-native American English (NNAE) speech. Lindemann 2005 provides evidence

for which NNAE accents are held as folk artifacts. She had American English respondents evaluate international English on world maps and also had them rate the English of 58 countries. Her results suggested that for many Americans, China represents Asia and the Far East, Mexico stands for Latin America, no single country represents (Western) Europe, and nobody talks about Africa. In other words, Americans may have broad folk artifacts classified as "China" and "Mexico" that encompass entire continents; there is no broad artifact for Western Europe, several of whose countries have their own folk artifact (particularly France, Germany, and Italy); and there are no artifacts whatsoever regarding Africa.

Lindemann also noted several prominent subgroups, which we may assume are those which have the strongest associated folk artifact. These subgroups included Chinese English, which was evaluated very negatively and likely represented all of Asia; somewhat-negatively evaluated Mexican English, which likely stood for all of Latin America; and "harsh and guttural" Russian English. Judgments of Indian English and German English, two other significant subgroups, were mixed. There were two different frameworks for both of these groups, one positive and one negative. For Indian English, there was a positive association with British English and a more negative association represented by "Kwicky Marts" from *The Simpsons* television show. For German English, there was a positive regional association with Western Europe and a more negative regional association with Russia. Though it wasn't a significant subgroup, another group with competing frameworks was French, which was considered both romantic and arrogant. Lindemann concluded that "sociopolitical factors and familiarity can largely explain

respondents' patterns of evaluation of the English of these countries, with countries that may be identified as adversaries of the US and less familiar countries being rated most negatively" (195). Her study demonstrates that extra-linguistic factors play a role in peoples' opinions about and awareness of specific folk artifacts.[4]

Awareness of folk artifacts can be reflected in listeners' ability to differentiate between various artifacts. Several studies have established that listeners can distinguish between AE and NNAE speech, starting with Bush 1967, who showed that listeners could consistently distinguish between American, British, and Indian English accents. This held not only for sentences (which were correctly distinguished 100% of the time), but also for nonsense words (91% correct). Munro 1995 found that listeners could distinguish AE and NNAE (Mandarin) speech even when presented with only nonsegmental information. Ikeno 2005 also demonstrated that AE listeners could accurately detect NNAE speakers as non-native, based on familiarity and "prototypical, conceptual representations of accents" – i.e. folk artifacts. Flege 1984 determined that listeners could correctly distinguish AE speech from NNAE (French) speech. While they performed best with long stretches of speech, listeners could make this distinction on the basis of hearing words, syllables, single phones, or even just parts of a phone (in this case, the burst of /t/).

---

[4] Although the creation and transmission of folk artifacts is outside the scope of this work, we may assume that an individual's personal experiences and the media are some extra-linguistic factors which influence this. Preston maintains that "one of the results of media exposure (though not a necessary one) is that some variety (or some aspects of it) may gain folk artifact status. I assume, however, as I believe most folklorists would, that such artifacts are more commonly passed on (and have a livelier and longer existence) through traditional (oral, face-to-face) cultural means" (59).

In addition to distinguishing between accents, awareness is reflected by listeners' ability to recognize specific accents. In a forced choice task, Derwing & Munro 1997 found that AE speakers could recognize NNAE accents at a rate better than chance (on average, 52% of the time). In their study, misidentification errors revealed that listeners most often confused the Cantonese and Japanese accents for each other and the Spanish and Polish accents for each other. The misidentification of Asian languages and European languages is revealing. It seems likely that listeners appeal to broad folk artifacts, including geographically-based ones, despite differences between the accents and even between the native languages. The results of Clopper & Pisoni 2004b, who found that listeners' accent categories can differ from categories linguists have established, supports this. Listeners were not as sensitive to a linguistic six-accent division of regional varieties of American English, instead preferring to lump speakers into three large groups. While researchers may divide a large group along distinct scientific lines, laypeople seem to have broader conceptual categories. Podbresky et al. 1990 affirm this is also the case for NNAE accents. Specifically, in their pilot study they found that listeners could not differentiate between four Hispanic accents (Cuban, Costa Rican, Argentinean, Puerto Rican) or between four Asian accents (Chinese, Japanese, Vietnamese, Korean). The lack of distinction between the Asian accents is even more surprising, given the fact that these languages belong to different language families. In sum, it appears that while some listeners may narrowly distinguish between similar folk artifacts, many others lump them together into one broad artifact.

The presence of broad folk artifacts, specifically an "Asian" artifact, finds confirmation in Lindemann 2003. In a free identification task, she had listeners evaluate AE speakers and Korean NNAE speakers and then asked what ethnicity each of the speakers was. Listeners only identified the NNAE speakers as Koreans 8% of the time, although they were identified as Asian, Chinese, or Japanese more than half the time (see Figure 2). Lindemann noted the importance of misidentification patterns, concluding that "stigmatized non-native accents are categorized together and are for the most part ethnically undifferentiated" (359). We may expand this to assert that, while folk artifacts do exist for specific language varieties, they also exist for broadly classified groups, even when those cross language, geographic, and ethnic lines.

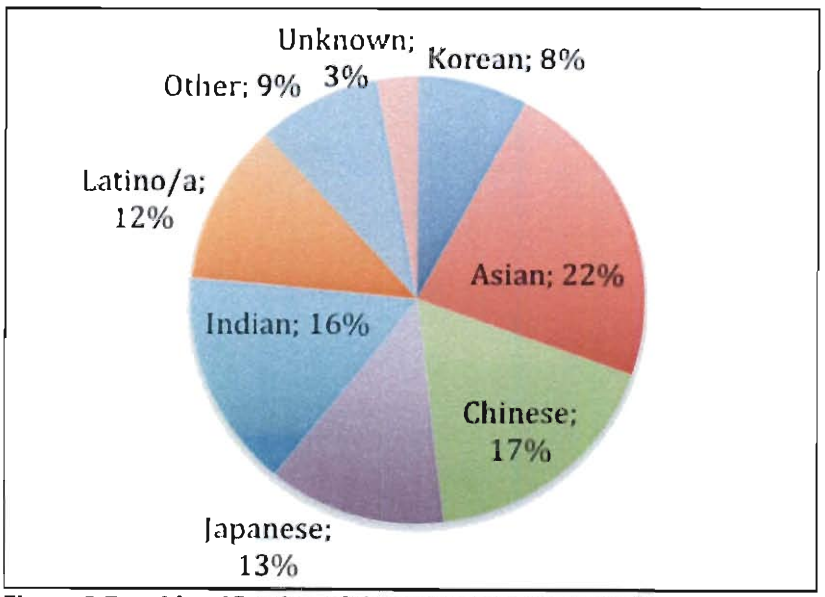


Figure 2 Free identification of Korean voices, from Lindemann 2003.

The ethnic undifferentiation referred to by Lindemann may be due in part to *the other-race effect*. This is a psychological phenomenon whereby people in one

racial or ethnic group have more difficulty distinguishing faces of another ethnic group than distinguishing faces of their own group, perhaps best represented in pop culture by the phrase "They all look alike to me." Chance & Goldstein 1996 comment that "the other-race effect is easily replicated and substantially affects subject accuracy in recognizing faces" (155), although it does diminish with increased contact between groups. The other-race effect is an obstacle to speech perception as well as facial perception. Kerstholt et al. 2006 refer to this in explaining listeners' poor performance for the recognition of an accented speaker in a voice lineup study.[5] Doty 1998 goes so far to conclude that the other-race effect is not limited to race, but also impacted by nationality. In his voice lineup study, participants were able to recognize individual speakers from their own countries significantly more often than those from other countries. This finding was not limited to non-native English speakers; it held even with native English speakers of the same race from Britain and America, as participants recognized native English speakers from their own country 88% of the time and native English speakers from another country only 16% of the time. It appears that race is not in fact the critical feature in recognition, but the deciding factor is whether a speaker is in-group or out-group.

It is widely accepted that speakers modify their speech based on their audience (Giles 1977; Bell 1984), and particularly whether or not the interlocutors

[5] A voice lineup study is a technique used in forensic linguistics where a listener hears a target speaker and later attempts to identify the same speaker from a group of several different speakers. (The target may be present or absent from this group.) This is similar to an eyewitness forensic task, except the participant is recalling what they have heard instead of what they have seen. For more details on this methodology, see Hollien 1996 and Yarmey 1995.

are in-group members (Tajfel 1978; Duszak 2002).[6] In addition to its effect on speech production, whether a speaker is perceived to be in-group or out-group can also affect listeners' perception (e.g., Niedzielski 1999, Kang & Rubin 2009). The variations in perception highlighted by the voice lineup studies above seem to show that once listeners have categorized a speaker as an out-group member, they are less likely to recognize that individual. The listener associates the out-group speaker with a folk artifact, and then seems to be less aware of the details of their speech. It may be that artifact association overwhelms their perception of specific details, or it may be that there is simply less need for awareness of details because they can be provided by the artifact. Either way, listeners immediately categorize a speaker as in-group or out-group and subsequently are more or less aware of the details of the individual's speech depending on the category. They clump together various out-group speakers according to folk artifacts, but use more fine-grained characteristics to distinguish in-group speakers.

This does not mean that listeners do not have any details associated with the folk artifacts. They simply rely on the folk artifacts to provide the details, more so than relying on the individual speaker. The folk artifact is a sort of cognitive

---

[6] It is also interesting to note that the status of an accent as in-group or out-group may vary, relative to the other accents present. For example, Abrams and Hogg 1987 found that native speakers of a Dundee (Scottish) accent preferred their own dialect to a Glasgow (Scottish) accent and RP (British) accent. However, when they were only presented with the Glasgow and RP accents, they preferred the Glasgow accent; it had become the in-group variety. This would be similar to an American listener from Boston preferring Boston accents to Southern US accents and Russian accents, but preferring Southern US to Russian when the Boston accent wasn't present. Lindemann 2005 suggests group relativity may also explain why her American participants evaluated the English speech of Asian speakers more negatively than Latin American speakers: "while all non-native speakers may be considered out-group in comparison to native speakers, when non-native varieties are considered together, Latin American speakers are in-group compared to the Asian speakers" (209). In cases where non-native speech is all that is available, certain non-native groups may be preferred to other non-native groups.

template the listener can access. The level of detail within the folk artifact can vary based on the strength of the artifact, along with the other modes of linguistic awareness. As suggested above, not all varieties have a strong folk artifact associated with them. However, as Preston 1996 explains, "Folk linguistic awareness is enhanced (perhaps particularly in the 'control' mode) when the object is a folk artifact, an object which may be featured in performance" (63). Strong folk artifacts have a high "performance potential." Preston reflects that much can be learned about folk linguistic awareness by investigating folk artifacts and particularly the details of imitations of artifacts (71). The present research pursues this by analyzing imitations of strong folk artifacts.

1.3 Imitation

Imitation is a principle type of linguistic performance, in which speakers present their interpretation of another variety or speaker.[7] Evans 2002 explains that imitation "refers to the conscious use of a variety which is not the speaker's usual vernacular" (96). Imitation is rooted in the speakers' perceptions of the other variety, and these perceptions may be based on experience with the variety and/or folk artifacts. It is reflected in the control mode of linguistic awareness.

Because of the sociolinguistic focus on vernacular as opposed to performed speech, little research has been done on imitations. Preston 1989 commented, "In general, though the suggestion has been around for some time, extensive collection

[7] While imitation has been a focus in other areas, such as language learning and language accommodation/convergence, this paper will not be addressing those types of imitation. The priority of this work is analysis of intentionally performed imitations.

of linguistic caricatures, particularly imitations of them in natural settings (e.g., narratives about speakers from other areas in which variety imitation is a feature of the performance), have not been extensively carried out" (13). Though there has still been no attempt to create an extensive collection of imitations in the twenty years since this quote, there are some studies which have explored this area.

In particular, the field of forensic linguistics has analyzed imitation because it can be used as a type of voice disguise. Forensic linguists typically employ a voice lineup methodology, where listeners hear a target speaker and then are instructed to identify the same speaker from a group of speakers. In perhaps the earliest work on accent imitation in voice disguise, Tate 1977 found that AE listeners from the South were able to correctly identify speakers of "General American" 88% of the time, authentic Southern dialects 76% of the time, and imitated Southern accents 62% of the time. She found no difference in identification between the imitated accents of untrained imposters and trained actors. Thompson 1987 found that AE listeners were worst at identifying people speaking Spanish in a lineup, best at identifying AE people speaking English, and intermediate at identifying AE speakers imitating Spanish-accented English. Markham 1999 found that individual speakers varied considerably in their success at imitating regional accents and even linguistically trained listeners varied considerably in their success at identifying the accents and their authenticity.

Another form of imitation which forensic linguists have analyzed is *impersonation*: mimicry of a specific individual. Schlichting & Sullivan 1997 found that high-quality impersonations can reduce voice lineup accuracy. Wretling &

Eriksson 1998 also found that impersonations were similar to the target voices. Zetterholm 2003 conducted a series of studies with Swedish impersonators in an attempt to discern which phonetic features were changed. She analyzed three impersonators mimicking the same voice, and found that while each of the imitations was different, all three impersonators chose to concentrate on the same prominent cues. This suggests that these impersonators were in agreement on the most important features of the voices they heard. Zetterholm determined that there were numerous important features, including pitch and intonation, voice quality, and dialect markers (segmental features), and that speech style and semantic content were also influential.

Forensic linguists have looked closer at some of these features within imitations and voice disguise. Lindsey & Hirson 1999 found that some speakers could change their non-standard production of /r/ when mimicking another accent, while others couldn't. Kunzel 2000 observed that speakers could consistently change their F0. Masthoff 1996 noted that several speakers with strong regional accents failed to conceal specific features of these accents when disguising their voices, despite being trained in phonetics and being allowed to choose their disguise method.

Imitation research outside of forensic linguistics has also shown that speakers can change a variety of features. Evans 2002 investigated a Southern imitation by a speaker who had lived in the South but not did have Southern features in his own speech. She found that he was able to modify his vowel system to mimic the Southern Shift. Further, listeners from the South rated him as Southern,

and almost all the listeners perceived his imitation as authentic. Flege & Hammond 1982 discovered that speakers familiar with a variety can actually imitate non-distinctive differences between languages (specifically, the difference in voice onset time of stops and final syllable lengthening between Spanish and English). Preston 1992 analyzed the cross-racial imitations of white and black AE speakers. White speakers used more features and seemed to have a larger repertoire in their imitations of black speakers than black speakers did of whites, and Preston observed that "one stands a better chance of discovering linguistic caricatures when the respondent group views the imitated group as a folk object" (352).

The folk artifact status of particular varieties can be highlighted by listeners' perceptions of imitation authenticity, and specifically whether listeners can distinguish between authentic and imitated accents. Neuhauser & Simpson 2007 had native German subjects listen to authentic non-native speakers and native German speakers imitating non-native (French and American English) accents. Listeners were instructed to state whether the accent was authentic or an imitation and to name the accent. Listeners were more successful at naming the imitated accents than the authentic accents. Additionally, listeners were not very successful at determining which accents were authentic and which were imitations. Although Neuhauser and Simpson do not present the specific numeric results or test them for significance, they "suggest that native German speakers and listeners seem to be in strong agreement about the stereotypical phonetic patterns which they consider characterise a particular foreign accent" (1808). In other words, the results reveal that speakers and listeners share powerful folk artifacts for these accents.

Torstensson, Eriksson & Sullivan 2004 also found strong agreement among their native Swedish speakers as to what constitutes a foreign (British) accent. When three speakers were instructed to spontaneously imitate the non-native accent, they all modified similar prominent phonetic features. Speakers were then exposed to an authentic accent and asked to again imitate this accent. After training with the authentic accent, further changes were made to the imitated vowels, consonants, and prosody. Torstensson et al. considered the shared features to compose a "cognitive prototype" of this accent. It is beneficial to break down their cognitive prototype concept by their two types of imitations. The first set of spontaneous imitations revealed the shared folk artifact. The second set of modeled imitations built on that folk artifact, adding features based on the listeners' perception of an individual's speech. The specific linguistic features were initially provided by a shared folk artifact, and none of these features were removed from the second imitation, even when the target did not use them. Based on this we may theorize that folk artifacts are the basis of imitation: perception of individuals can build on this, but it is difficult to change the foundation. This leads to the question: are the phonetic features that contribute to the imitation the most salient features of the folk artifact?

1.3.1 Salience and Imitation

Salience is a foundational concept in sociolinguistics, though it is infrequently discussed and even more rarely quantified. Torbert 2004 asserts, "Every sociolinguist knows what *salience* means and possesses some notion of which

linguistic variables are highly salient, but such notions remain under-investigated and largely un-reflected on" (2). In fact, Torbert 2004 and Hickey 2000 are perhaps the only two linguistic works specifically focused on salience. Hickey defines salience as "a reference to the degree to which speakers are aware of some linguistic feature" (57). Torbert's definition is "the degree to which the language-speaking public are aware of specific correlations between linguistic variants and external social characteristics" (1). Both of these understandings of salience center on speaker awareness. We can use the modes of folk linguistic awareness to create a more complete definition: salience is the availability of certain details of a language variety.

Salience is often associated with importance. However, many linguistic features may be salient, and determining the relative importance of these features is quite difficult (Thomas & Reaser 2004). Hickey suggests that salience can be prompted by a variety of triggers, including acoustic prominence, the merging of two phonemes, linguistic unconformity, deletion and insertion, and the retention of conditional rules. Trudgill 1986 contends that surface phonemic contrast and degree of phonetic difference are two of the most important factors for the salience of accent features. He continues, "Other factors presumably remain to be detected, but in any case the salience of features can often be determined by an examination of the process of imitation" (37).

The idea that imitation reflects salience is a common one. Trudgill also states, "Obviously the most salient features... are precisely those which are reproduced during *imitation*" (12). Hickey agrees, "It is probably fair to say that the elements in

a variety or language which are most salient for its speakers are those used in linguistic stereotypes" (58), where by "linguistic stereotypes" he is referring to imitations of folk artifacts. Preston 1992 comments, "Perhaps the most striking, overt proof of the salience of speech differences to nonlinguists occurs when they imitate another variety" (327). Wertheim 2003 claims that performance speech can "demonstrate which elements of a language or dialect are most salient to its speakers" (520). Finally, Schilling-Estes 1995 asserts, "When speakers attempt to "put on" a dialect for an audience, they enhance what they perceive to be the salient features of that dialect variety. Thus, through examining performance speech, we can gain insight into which aspects of linguistic production are most salient to the performer and his or her audience" (126).

The connection between imitation and salience is merited. Imitation is rooted in the speaker's perception of another speaker or variety (or folk artifact of a variety) and reflects their awareness of it. However, linguists must be careful to differentiate the modes of linguistic awareness, and not assume that imitation (the mode of control) is identical to the other modes. Recall that our definition of salience involves the modes of availability and detail; it does not include control or accuracy. While high control certainly can reflect salience, the correlation is not necessarily one-to-one. This leads us to aspects of imitation about which we must be cautious.

1.3.2 Complexities of Imitation Research

There are three issues that complicate imitation research. These are a conflation of the modes of linguistic awareness, an unclear perceptual target, and the interpretation of details.

The first difficulty of imitation research is that the modes of linguistic awareness are not usually distinguished. Yet as discussed above, control does not necessarily exactly correspond to the other modes of linguistic awareness. For example, while the features modified in an imitation may indeed be more available, they may be inaccurately modified. Preston warned that when it comes to performances, "a successful imitation need be neither complete nor accurate" (1996:65). He notes that it is "difficult to decide when folk imitations are 'inaccurate' due to lack of knowledge and/or ability and when they are inaccurate because folk performers feel that they have done enough to effectively establish whatever point is being made by the imitation" (66). Even when an imitated feature is inaccurate, the fact that it is being modified may indicate that the feature is highly available. There may be other highly available features that a speaker is unable to control, and so go completely unimitated. While imitation can reflect salience, there is not a simple correlation among the modes of control, availability, detail, and accuracy. However, as long as the investigator recognizes and distinguishes between the modes of linguistic awareness, they can avoid the conflation misstep. In my study, this is done by concluding that the features which are modified in imitations are controllable and likely more available; however, features which are not modified are not necessarily assumed to be less controllable (they may be controllable but

not differ between the imitation and the regular voice), nor are these features assumed to be less available (they might be very available but completely uncontrollable). Additionally, accuracy and detail are not assumed to be revealed by imitation modifications alone.

The second issue is that the exact target of an imitation is often unclear. Imitation is perhaps the best way linguists can get at speakers' perceptions of a target from a production standpoint. Niedzielski and Preston 2000 affirm, "Mimicry appears to be the most productive means of eliciting the details of varieties from the folk" (111). This is because while people usually cannot explicitly discuss the details of another variety, they can more often produce them in imitation (Preston 1996:45). However, the difficulty here is that the details of an imitation may be less reflective of listeners' perception of an actual reference, and more reflective of a folk artifact of that variety. Listeners' expectations powerfully affect their perceptions (Strand 1999, Johnson et al. 1999, Niedzielski 1999, Drager 2006, Hay et al. 2006, Koops et al. 2008), and if listeners have categorized a speaker or variety as a member of a folk artifact, their imitation may be based more on their perception of the folk artifact than of the actual target. This difficulty may be relieved by acknowledging the likelihood of the target being a folk artifact, and clearly recognizing what type of perception is being assessed. In the present research, the two different targets (folk artifacts and immediate authentic accents) are distinguished by eliciting two different types of imitations.

Finally, the third complexity of imitation research is in interpreting the specific linguistic modifications in an imitation. Preston 1992 listed some of the reasons this is difficult:

> "1) In variety imitations, the majority of features used are shared by many varieties and may even be present in the imitator's variety.
> 2) Some of the features which might qualify as caricatures do so only on the basis of their frequency of use in the imitated variety and actually occur, albeit at different levels of frequency, in the variety of the imitator and/or in a number of other varieties.
> 3) Although it seems obvious that indicators would not be used in imitating the speech of others, it is not at all clear that caricatures are exclusively a sub-class of stereotypes; respondents who imitate other varieties rather effectively are not able to comment at all on the features they have used in the performance.
> 4) Imitations may contain caricatures which are not features of the variety being imitated at all" (328).

A linguist must keep these caveats in mind when analyzing the details of imitations. Thorough knowledge of both the speaker's original variety and their target variety will also help reduce the risk of misrepresenting certain features. My study achieves this by comparing the imitations to both the speakers' "regular" voices and to the authentic voices which serve as targets.

Research involving imitations must be mindful of these three potentially complicating factors: conflating the modes of linguistic awareness, being unclear on the perceptual target, and misinterpreting the linguistic details. Acknowledging these issues goes a long way towards mitigating them, and grounding imitations in

the concepts of folk artifacts and folk linguistic awareness also alleviates the difficulties.

1.4 Summary

This chapter introduced the concepts of linguistic awareness, folk artifacts, and imitation. I started with a history of awareness in sociolinguistics, introducing Labov's concepts of social awareness and attention to speech and displaying that performance speech can be regular despite speakers' high awareness, concluding that Preston 1996's folk linguistic awareness is the most comprehensive framework. This model distinguishes four aspects of awareness: availability, detail, accuracy, and control. Folk artifacts both reflect and reinforce this awareness, and they may be specific or broad. Imitation, a type of performance speech, can reveal linguistic awareness and display folk artifacts. It is represented in linguistic awareness by the control mode, which is different from the modes of availability and detail that express salience. While imitation research may be complex, care on the part of the researcher can mitigate the difficulties.

The present research addresses the three concepts of awareness, folk artifacts, and imitation by analyzing speakers' imitations of NNAE speech and listeners' perceptions of them. The speakers' initial imitations reflect their awareness and the folk artifacts they hold; subsequent imitations reflect their perception of actual NNAE speakers. Listeners' evaluations reflect their awareness and folk artifacts, as they categorize both imitated and authentic NNAE accents.

The outline of the rest of this dissertation is as follows. The second chapter provides further background for this research. It describes a pilot study which explored listeners' recognition of various NNAE accents, and how that study influenced the present work. Hypotheses and methodological specifics are explained in the third chapter. The fourth and fifth chapters present the results from the speakers' imitations and listeners' evaluations. The final chapter discusses the results and draws conclusions from this research.

Chapter 2 Can Listeners Identify Accents? A Pilot Study

This chapter presents my preliminary work in the field of foreign accents and how that influenced methodological aspects of the present research. I conducted a free-response survey where American English (AE) listeners attempted to identify the home country of the non-native speakers of American English (NNAE) they heard. The identification of accents revealed broad folk linguistic artifacts, which are more likely based on geographic region than specific country or language.

2.1 Accent Recognition Survey

In my pilot study (Brunner 2008), I investigated the recognition of NNAE accents by AE listeners. More specifically, I was interested in listeners' free responses to NNAE speech because it would be revealing of listeners' perceptual classifications.

This research falls under the realm of perceptual dialectology. Most perceptual dialectology research has involved regional accents. Examples of this are Clopper's research in the United States (e.g., Clopper & Pisoni 2004a; Clopper & Pisoni 2004b; Clopper & Pisoni 2006) and research compiled by Preston from both within the US and abroad (see e.g. Preston 1999; Long and Preston 2002). Linguists have also investigated the perception of ethnic varieties (e.g. Purnell, Idsardi & Baugh 1999;

Thomas & Reaser 2004). While regional and ethnic accent recognition has been addressed, much less research has been conducted on identification of NNAE accents. As discussed in section 1.2, listeners can generally distinguish NNAE speech from AE speech when they hear it (Bush 1967; Flege 1984; Munro 1995; Ikeno 2005) and recognize NNAE accents in forced choice tasks (Derwing & Munro 1997). However, there has been no research conducted on listeners' free response identification of NNAE accented speech. This type of identification is important, because it reveals listeners' perceptual categorizations without being influenced by the researcher's classification system. Clopper & Pisoni 2007 conclude that "labels provided by the experimenter in a forced-choice task may lead to some response biases that can be reduced by using a free classification task" (436). The only free response perceptual dialectology research that has been conducted on AE listeners with NNAE speakers was a study by Lindemann 2005 (discussed extensively in 1.2); however, in this study participants were not exposed to actual NNAE speech. This leaves open the question of how listeners would perform in free response identification of actual NNAE speech. This pilot study sought to answer this.

In addition to analyzing listeners' free identifications, I was also interested in the influence of previous listener interaction with NNAE accents on their identification accuracy. Derwing & Munro 1997 found that greater amounts of previous listener contact with NNAE accents improved identification. Several other studies have also found that listeners' regional accent identification improves with greater familiarity, where familiarity is based on listeners' geographic location and

mobility (e.g., Clopper & Pisoni 2004a, 2006; Williams, Garrett & Coupland 1999; Baker et al. 2009).

Therefore, this pilot study addressed two research questions:

1. How accurate are listeners at identifying NNAE accents in a free response task?
2. How do listeners' own interactions with NNAE speakers affect their judgments?

To investigate this, AE respondents heard voices of NNAE speakers from seven different countries and were asked to identify the speakers' countries of origin. After this, listeners were asked about the amount of interaction they had with NNAE speakers. The following section describes the specific details of this study.

2.2 Methodology

Seven NNAE speakers were recorded reading the same short passage (approximately 45 seconds in duration). These speakers self-identified as Australian, Korean, Finnish, Portuguese, Turkish, Mexican, and Palestinian. A variety of speakers from different countries were chosen to take part, instead of several speakers from the same country. While the latter method would have lessened speaker-specific effects (including important variables such as age, gender, and accentedness), the former method was chosen to get a broad sampling of different accents for this pilot study. Listeners heard each NNAE speaker and were then asked what country they thought the speaker was from. (As suggested above, this

question was purposely open-ended so as to not lead the listener. However, this free response design prohibited a statistical analysis of the data.) Listeners then heard six more speakers and were asked the same question for each of them. Listeners concluded by providing basic demographic data about themselves.

The survey was conducted on the Internet. This facilitated obtaining a large number of responses in a short amount of time. Participants were recruited through email chains and Facebook (a social networking website), allowing numerous people from different parts of the United States to complete the survey. Entitled *Name the Accent!*, this study was billed as a fun survey, similar to many other online surveys that people freely participate in. A total of one hundred and fifty-five AE speakers completed the survey.

2.3 Results

Very few listeners accurately identified the countries speakers were from (see Figure 3). The most easily recognized speaker was from Australia, with 43% of listeners correctly identifying her country of origin. The next most accurate identification was for the Mexican speaker, with 19% of listeners correctly recognizing her. Seven percent of listeners correctly identified where the Korean speaker was from. Only 3% of listeners correctly identified where the Turkish speaker was from and 2% correctly identified the Portuguese speaker. No one

correctly identified where the remaining two speakers were from (Finland and Palestine[8]). The average rate of correct identification was 11%.

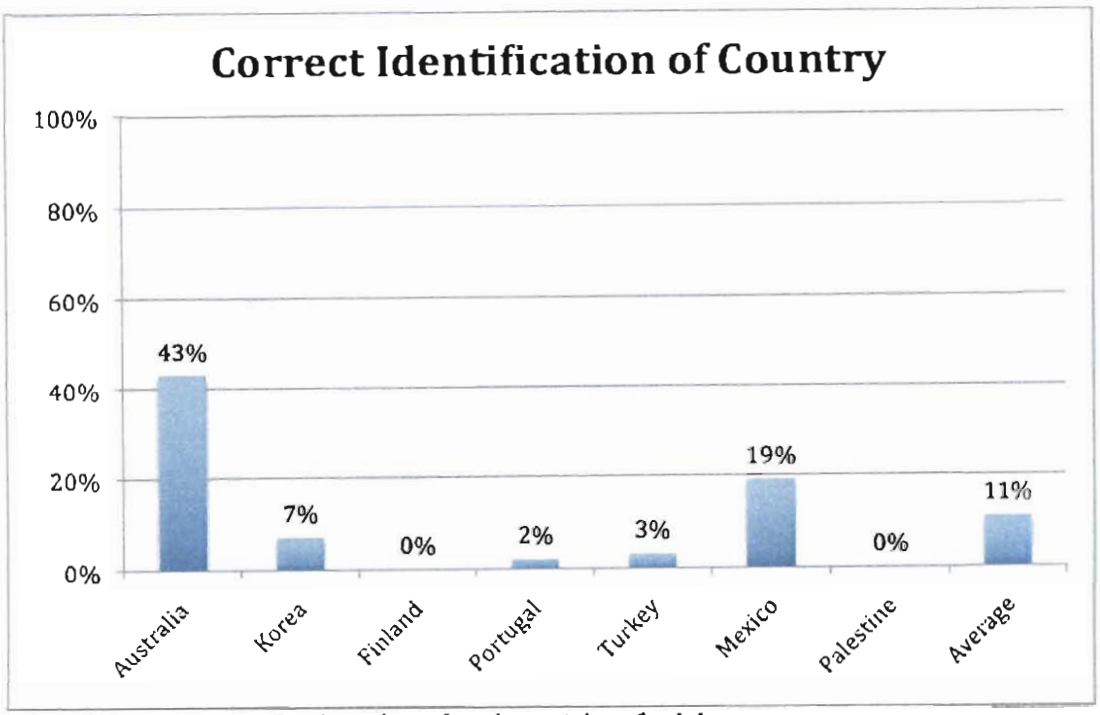


Figure 3 Correct identification of speakers' countries of origin.

For some speakers, there were repeated incorrect identifications. The most frequent incorrect responses for the speaker from Australia were English-speaking countries (77 responses), including the United Kingdom (45 responses) and New Zealand (20 responses). In fact, the top three country responses for the Australian speaker accounted for 85% of the responses (see Table 1). This is quite different from the Finnish speaker, who was never correctly identified. The top three responses for the Finnish speaker only accounted for 37% of the data and total responses spanned a range of 39 countries. With more than twice as many countries

[8] The Palestinian speaker was not identified as Palestinian nor as being from Jordan, the West Bank, Gaza, or Israel.

and less than half as much data accounted for, this suggests that the responses were much less consistent for the Finnish speaker than for the Australian speaker. With the exception of Mexico, the countries of the remaining speakers did not even make it to the top three responses; for example, the Korean speaker was most frequently thought to be from China (29 responses) and Japan (22 responses). Based on these findings, I decided it would be constructive to analyze response accuracy based on the region of the world speakers were from.

Speaker	Number of Countries Guessed	Top 3 Responses (% responses for each country)			Percent of Total Responses
Australia	15	.43 Australia	.29 UK	.13 New Zealand	.85
Korea	38	.19 China	.14 Japan	.08 Spain	.41
Finland	39	.16 France	.11 China	.10 Russia	.37
Portugal	37	.23 Russia	.11 Germany	.06 France	.41
Turkey	30	.16 Russia	.15 China	.09 Japan	.41
Mexico	40	.17 Mexico	.13 Spain	.06 India	.37
Palestine	31	.19 India	.11 Ireland	.10 Germany	.40

Table 1 Most popular listener responses. The second column shows the number of different countries guessed in the responses. The next section shows the three most popular responses, and the percent of responses that selected that country. The final column shows how much of the total responses for each speaker were accounted for by the top three responses.

Since respondents were only asked to identify the country that speakers were from, the region of the world speakers were from had to be extrapolated from these answers. This approach assumes that the respondents would be able to correctly state where in the world the country they answered is located. According to the National Geographic-Roper Survey of Geographic Literacy (2006:20), only 15% of 18-24 year old Americans are able to correctly place nine major countries

and natural landmarks with their correct continents, so this is in fact a very large assumption. The correct region identification rate should therefore not be taken as a true result.

The world was split into seven regions for this task: North America, Latin America, Australia, Africa, Europe, Asia, and the Middle East. A difficult country to classify was Turkey, as the perceptual line between the Middle East, Europe, and Asia appears to be a somewhat fuzzy boundary. Due to cultural and geographic similarities, Turkey was classified as part of the Middle East for this study.

Accuracy rates improved tremendously when analyzed by region instead of by specific country. The overall correct identifications jumped from 11% to 39% (see Figure 4). Five of the speakers' regions were correctly identified more than half the time: the Portuguese speaker was identified as European 71% of the time; the Australian speaker was identified as Australian 55% of the time; the Korean speaker was identified as Asian 53% of the time; and the Finnish speaker was identified as European 51% of the time. The Mexican speaker was identified as Latin American 33% of the time. The Palestinian speaker was identified as Middle Eastern only 3% of the time, and the Turkish speaker was identified as Middle Eastern 6% of the time.[9]

[9] Had Turkey been classified as a European country that rate would have increased to 39% (raising the overall correct identification to 44%); had Turkey been classified as an Asian country the rate would have increased to 47% (raising the overall correct identification to 45%). Leaving Turkey out of the calculation would result in a correct identification rate of 44%.

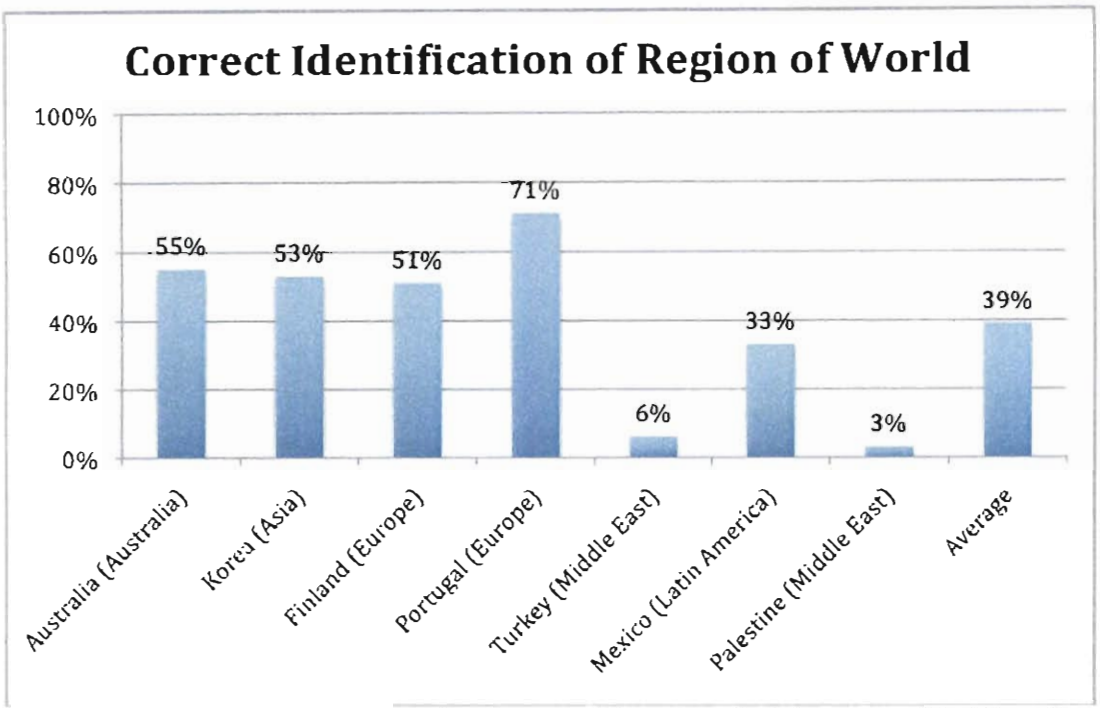


Figure 4 Correct identification of speakers' geographic region of origin, extrapolated from the country responses.

The lowness of correct region identifications for countries such as Mexico could be partially explained by the amount of people who answered Spain, a Spanish speaking country which is not in Latin America. This led to an analysis of the language of all responses. Country responses were recoded by the dominant language in the country. Once again, caution must be exercised when extrapolating from the responses, as survey responses of languages would probably differ from the responses of countries. Indeed, it is unlikely that respondents would be able to list the languages of all the countries included in the responses.

Forty-five languages were represented in the responses. We see a leap to 92% correct identification of the Australian speaker as an English speaker (see Figure 5). There is also improvement of the Mexican speaker's identification as a

Spanish speaker (43%). While this method of coding improved results for the Australian and Mexican speakers, it drastically reduced accuracy for the remaining countries. The Korean and Portuguese speakers were only identified 7% and 5% of the time, respectively. The Turkish and Palestinian-Arabic speaker were identified at just 3% and 2%, respectively. No one recognized the Finnish speaker. The overall correct identification was 22% when analyzing by language.

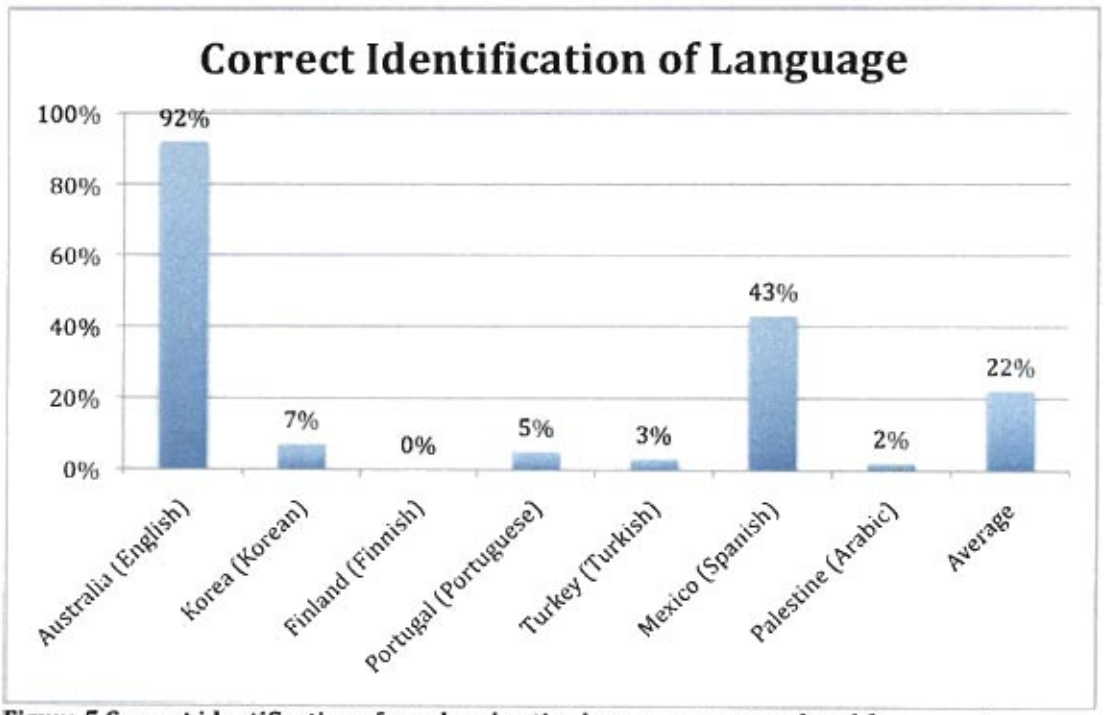


Figure 5 Correct identification of speakers' native languages, extrapolated from country responses.

Several responses for the Portuguese speaker included other Romance languages, such as Spanish and French. Because of this, a final extrapolated analysis was done to assess responses by language family. Fourteen language families were represented in the answers: Germanic, Italic, Balto-Slavic, Indo-Iranian, Altaic, Semitic, Sino-Tibetan, Austronesian, Tai-Kadai, Austro-Asiatic, Afro-Asiatic, Hellenic,

South Caucasian, and Finno-Ugric. While most respondents would not be familiar with these language families or be able to identify speakers using this classification system, this evaluation was done for the sake of thoroughness.

Responses proved more accurate when the data was classified in this manner than by specific language, with overall correct identifications at 32%. In looking at Figure 6, it is obvious that the correct identifications for language and language family are again due to the higher scores for the Australian and Mexican speakers; the identifications were not as strong for the other speakers. Almost all responses correctly identified a Germanic language country for the Australian speaker (94%). Responses were also fairly consistent in identifying Italic language countries for the Mexican speaker (60%). The language family for the Portuguese speaker (Italic) was correctly identified 25% of the time. The Turkish speaker and Korean speaker were both identified as Altaic languages 21% of the time.[10] Finally, the language families of the Palestinian speaker (Semitic) and the Finnish speaker (Finno-Ugric) were identified only 2% and 1% of the time, respectively.

[10] The Korean language is generally considered to be an isolate. However, for this exercise it was classified as Altaic in an effort to improve the identification results.

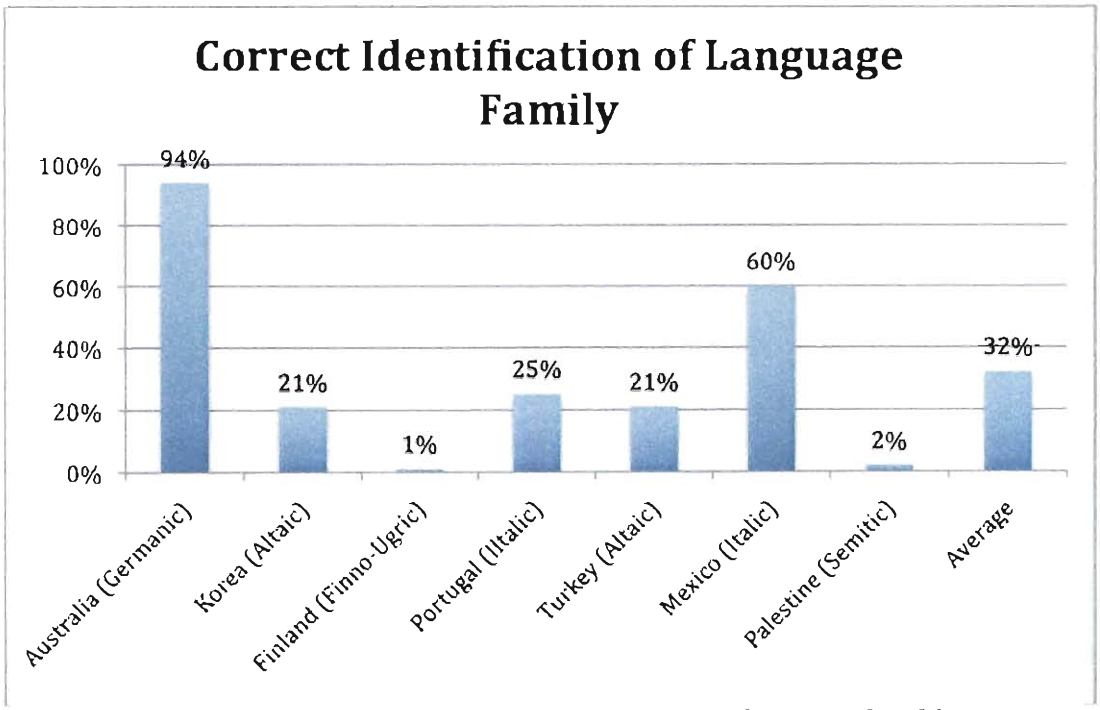


Figure 6 Correct identification of speakers' native language family, extrapolated from country responses.

Overall, we see that listeners are best at identifying speakers by their geographic region, averaging 39% correct when doing so. Listeners are likely to be more successful at this task than at identifying speakers by their country of origin (11% correct), native language (22%), or native language family (32%). From this we may hypothesize that listeners categorize speakers more by region than by country or language. This conclusion is fairly speculative, as it is based on extrapolated results instead of original data, and further research is necessary to confirm this.

2.4 The Role of Interaction

At the end of the survey, listeners were asked, "*How often do you interact with people who are non-native English speakers? Where are they from? (For example, if you have a Puerto Rican coworker that you see daily or an Indian doctor you visit every couple of months.)*" The question was expanded after approximately 50 listeners had answered, to ask for more detail in where these interactions took place and the length of these interactions. The answers varied widely, from some listeners answering "none" or "infrequently" to others answering "daily" and listing numerous countries or ethnicities.

When listeners included at least one country in response to this question, their responses were analyzed to see how that affected their answers for specific countries. There were four possible scenarios for each NNAE accent that the listeners claimed experience with (see Table 2). The first scenario was one in which a listener interacted with a speaker from another country and accurately identified the study speaker from that country. This scenario was the most rare. Very few of the responses (16 out of a total 1042 responses, less than 2%) were accurate identifications of a speaker from a country the listener asserted familiarity with.

	Accuracy	Scenario	Example	Percent of Responses
1	Correct	Listener lists a country and correctly identifies a speaker from that country.	Listener claims regular interaction with Spanish speakers from Mexico. They accurately identify the Mexican speaker.	2%
2	Incorrect	Listener lists a country and incorrectly identifies the speaker from that country as from a different country.	Listener claims daily interaction with a Korean speaker. They misidentify the Korean speaker as Serbian.	3%
3	Incorrect	Listener lists a country and incorrectly identifies a speaker from a different country as from that country.	Listener claims daily interaction with a German speaker. They misidentify the Portuguese speaker as German.	5%
4	Unknown	Listener lists a country but does not identify any speakers as from that country. OR Listener does not list any countries.	Listener claims weekly interaction with Vietnamese speakers. They do not identify any of the speakers as Vietnamese.	90%

Table 2 Possible scenarios of listener interaction and listener accuracy.

The second scenario occurred when the listener interacted with a NNAE speaker from one country but then did not correctly identify a study speaker who was from that country. This happened in 30 instances (3% of all responses). For example, while four listeners said they interact with people from Korea on a daily or weekly basis, they misidentified the Korean speaker in this study as Philippine, Spanish, Serbian, and Columbian.

In the third scenario the listener cited experience with a NNAE speaker and then went on to use that speaker's country to incorrectly label one of the study speakers. This third scenario occurred in 53 responses (5% of all responses). For example, one listener said they interact daily with a Polish speaker. They then misidentified the Korean speaker in this study as being from Poland. Another

listener said they have traveled to Sweden; they went on to label both the Finnish and Portuguese speaker in this study as Swedish.

The fourth scenario occurred when a listener either asserted interaction with a NNAE speaker but did not label any of the speakers in the study as being from that speaker's country, or did not list any NNAE interaction. Accuracy could not be assessed in these instances. The vast majority of responses fell into this category.

Overall, most responses were not affected by other NNAE interaction. Ninety percent of responses did not list relevant experience with NNAE speakers; these speakers were therefore basing their responses on folk artifacts. While 10% of these responses with no known interaction were correct (the fourth scenario), almost 20% of those responses with interaction (scenarios one through three) were correct. Therefore, it appears that familiarity may slightly improve accuracy.

2.5 Conclusions

Listeners were very poor at identifying specific countries of origin of these specific NNAE speakers, correctly identifying the countries on average 11% of the time. When results were extrapolated to represent geographic regions, identification improved to 39%. Extrapolated results of native language averaged 22%, and native language family averaged 32%.

The effects of speaker interaction suggest that experience with a NNAE accent may slightly improve identification. Experience with other NNAE speakers occurs in ten percent of the responses, and 80% of those responses misidentify

speakers' countries; this is an improvement over the 90% misidentification rate of those listeners who had no interaction with these accents.

The correct identifications were usually highest for the Australian and Mexican speakers, and this is likely due to listeners' greater familiarity with these countries. Note that general familiarity is different from reported interaction. For example, despite their overall poor showing in the National Geographic-Roper study (2006:26), 88% of participants were able to correctly locate Mexico on a world map, and 74% were able to locate Australia. This indicates a general familiarity with these locations, regardless of whether participants have interacted with a speaker from there. This greater familiarity likely correlates with stronger folk artifacts for these accents.

While we may attribute the accurate responses for the Australian and Mexican speakers to strong folk artifacts, on the whole the free response identifications show us the relative inaccuracy of listeners' perceptions of NNAE speech. The extrapolated responses indicate that listeners are more likely to categorize speakers by geographic region than by country or language. It seems listeners may be more accurate when identifying broad (regional) artifacts than specific (national) artifacts.

2.6 Implications and Folk Artifacts

There were five limitations to this pilot study, each resulting in a suggested direction for the present dissertation research. First, one weakness of this pilot

study is that results may be influenced by speaker-specific idiosyncrasies. More speakers of each language would prevent this. Second, while free response questioning was appropriate for an initial investigation of categorization, it was not conducive to statistical testing. Forced choice responses are necessary so that quantitative results may be obtained (although c.f. the free categorization tasks in Clopper & Pisoni 2007; Clopper 2008). Third, the interaction responses varied wildly; a more objective way of measuring familiarity and amount of interaction is needed. Fourth, listeners had no way to signal their confidence in their choices, but some included question marks to indicate uncertainty. Another interesting and potentially revealing line of investigation to include would be that of listener's confidence (although this has produced mixed results, e.g. Hammersley & Read 1996). Fifth, some of the speakers' countries were too remote for listeners to identify. The cultural familiarity of Americans with other countries and language varieties must be taken into account; specifically, studies such as this must consider folk artifacts. Listeners cannot identify an accent if they do not have a folk artifact for it.

One way to determine which varieties do have folk artifacts is to analyze the most popular responses of this original pilot study. Listeners tended to identify some countries numerous times (such as China), while other countries (such as Laos) were only mentioned once. The top country responses in this survey are shown in Figure 7.

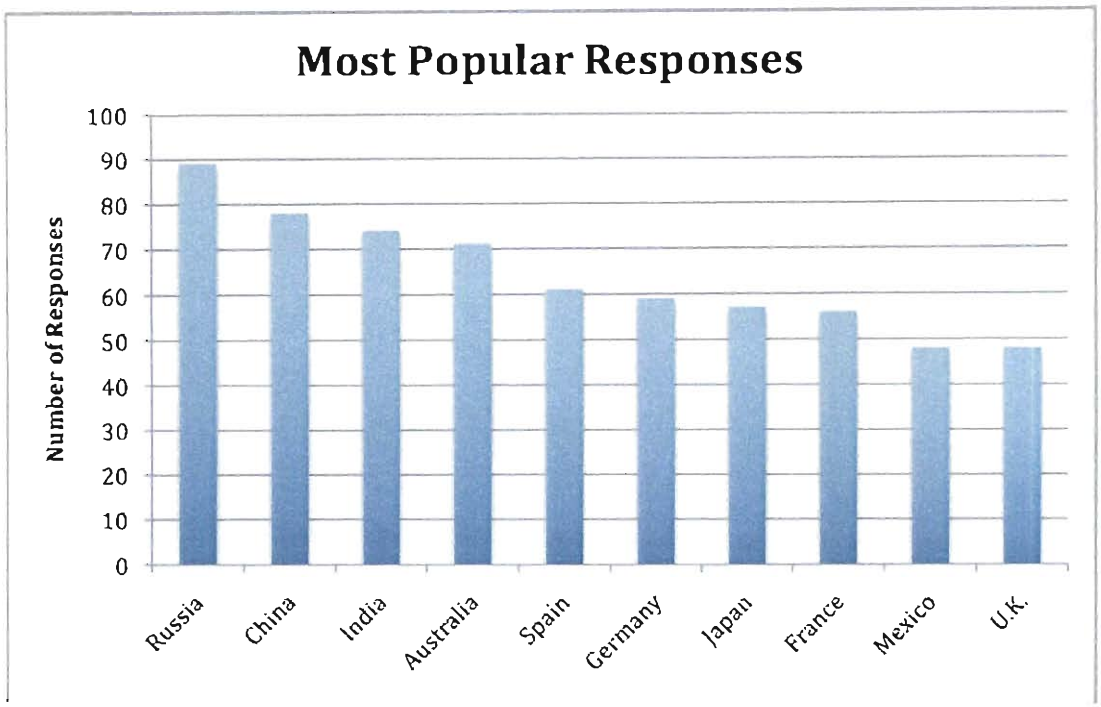


Figure 7 Most popular country responses, based on being included in more than 30 responses.

Russia was the most popular response, being guessed a total of 89 times. China, India and Australia followed, with 78, 74, and 71 responses, respectively. Spain (61), Germany (59), Japan (57), and France (56) came in next. Finally, Mexico and the United Kingdom, both with 48 responses, rounded off the countries which were included in more than 30 responses. Because these countries were named so frequently, we may assume that people are more familiar with these countries and more likely to have folk artifacts for them.

This is supported by Lindemann 2005, who found that the countries whose English was most commented on were China, Australia, Mexico, Russia, the UK, Canada, France, India, Germany, and Italy. Similarly, Lindemann's respondents indicated they were most familiar with the English of the following countries: US, Canada, UK, Mexico, Australia, Jamaica, France, China, Germany, Ireland, Japan, Italy,

India, Spain, Russia. The complementarity of the results from Lindemann and this pilot suggest that listeners are most familiar with these countries and most likely to have folk artifacts for them, and it is possible that listeners would perform better if they were exposed to speakers from these countries.

This dissertation therefore investigated the identification of several of the NNAE accents which are most strongly held as folk artifacts. A detailed methodology of this study, which implements the above suggestions, is presented in the next chapter.

Chapter 3 Methodology for a Study of Folk Artifact Imitation and Identification

In the previous chapter, we saw that listeners were fairly poor at identifying speaker origins based on their accents, and I suggested that listeners may be better at identifying them if the accents are held as folk artifacts. This brings forward the question of what folk artifacts consist of, and whether they can be successfully reproduced in imitations. To answer this question, I devised an experiment to analyze imitations of folk artifacts and listeners' identification of them. In this chapter, I begin by showing how this investigation builds on two previous studies of imitation. I then present the specific hypotheses which are tested. Finally, I detail the methods used, by describing the participants, the procedure they followed, and the process of phonetic analysis.

3.1 Motivation

While the pilot study discussed in Chapter 2 answered the basic question of how accurate listeners are at identifying NNAE accents, it also opened up many other questions. Would listeners be better at identifying accents for which they had a folk artifact? Could speakers reproduce these folk artifacts in imitation? Could

listeners recognize those imitated accents? Could listeners distinguish between authentic and imitated accents?

These questions were partially addressed in studies such as Torstensson et al. 2004, who analyzed speakers' foreign accent imitations, and Neuhauser & Simpson 2007, who analyzed listeners' responses to imitated and authentic foreign accents (c.f. 1.3) This research follows Torstensson et al. in eliciting two types of imitations from speakers, one before and one after hearing an authentic target speaker; it follows Neuhauser & Simpson in eliciting listeners' identifications of accent and authenticity of both imitated and authentic foreign accented voices. This investigation also expands on these studies in several ways. First, this research is tested for statistical significance to verify whether the results are meaningful, a step which the other studies did not take. Second, this study focuses specifically on American listeners, who are likely to have different types and degrees of foreign accent folk artifacts than the Europeans in the other studies. Third, this research looks at a greater number of accents at once, to facilitate comparison across a variety of artifacts. Finally, unlike the previous two studies, this research is grounded in the theoretical concepts of linguistic awareness and folk artifacts. Overall, the present study extracts the valuable techniques from the experiments by Torstensson et al. and Neuhauser & Simpson, developing a methodology that combines and expands on them.

Specifically, this research consists of a two-fold experiment investigating imitation and folk artifacts. In the first task, two types of imitations are elicited from AE speakers: free and modeled. The *free imitations* are spontaneous imitations of

NNAE accents which are presumed to reflect prevailing folk artifacts. The *modeled imitations* are performed immediately after hearing authentic NNAE accents, and are presumed to use actual speech perception to modify the folk artifacts. By comparing phonetic features across three guises (regular, free, modeled) for individual AE speakers, and comparing these to authentic NNAE speakers, we may assess the relationships between imitations, folk artifacts, and authentic accents.

The second task is that of listener identification: it tests whether listeners can first identify NNAE accents which have strong folk artifacts, and then classify them as authentic or imitations. Listeners' perceptions of the imitations will reveal the strength of folk artifacts. By analyzing awareness from the angles of speakers' imitations and listeners' identifications, we can begin to learn what imitation reveals about perception and how AE listeners perceive NNAE accents.

3.2 Hypotheses

The objective of this research is to explore the relationship between perception, folk artifacts, and imitations, from a linguistic awareness standpoint. More specifically, this research explores speaker perception as revealed by imitations of NNAE speech and listener perception as revealed by evaluations of imitated and authentic NNAE speech. In order to examine these relationships, the following hypotheses are tested.

1. **Speakers' modeled imitations will be more similar to the authentic accents than their free imitations.** If free imitations are based on folk artifacts, and modeled imitations expand on these artifacts with features heard in authentic accents, then the phonetic details of the modeled imitations will be more similar to the authentic accents. Torstensson et al. 2004 found this, as their participants added additional authentic features to their imitations after hearing an authentic speaker.
2. **Listeners will be able to identify the NNAE accents, both authentic and imitated.** If the accents being used are folk artifacts, then listeners will be able to recognize them. Neuhauser & Simpson 2007 found that their listeners were generally able to identify non-native accents, with the imitated accents being identified better than the authentic accents. My pilot study indicated that listeners are better able to identify accents which are held as folk artifacts (c.f. 2.6).
3. **Listeners will not be able to distinguish between authentic and imitated accents.** If the strength of the folk artifact association overpowers individual speaker cues, then listeners will not be able to differentiate between authentic and imitated accents. Neuhauser & Simpson found that listeners were unable to tell if an accent was authentic or imitated. The voice lineup studies discussed in 1.2 suggest that listeners are less likely to be aware of the details of an individual's speech once they have categorized them as an out-group member.

4. **Listeners will react differently to free imitations and modeled imitations.** If there are distinctive differences between the two types of imitations, then listeners will judge them differently. Modifications between speakers' free and modeled imitations will be reflected in listeners' perceptions of those imitations.

A detailed methodology for testing these hypotheses is presented next.

3.3 Experiment

This research involved three groups of participants: NNAE speakers, AE speakers, and AE listeners. This section describes these participants, the procedures they followed, and the method of phonetic analysis.

3.3.1 NNAE Speakers

The first group of speakers consisted of non-native speakers of American English (NNAE) who were recorded reading a short passage. Using a reading passage facilitated a narrow focus on the phonetic features of NNAE accents. To emphasize the performance aspect of this task, and therefore enhance the imitations (Preston 1996:64), the passage was a mock extortion threat (see Appendix A). This type of material is not uncommonly used in forensic linguistics research (e.g. Kunzel 2000). The recording took place in a university sound booth, and the entire

procedure lasted less than 15 minutes. Speakers were compensated with a $5 coffee house gift card.

NNAE speakers were primarily recruited from a university list-serv for international students. The objective was to obtain a minimum of two speakers (one male, one female) of each of the following languages: Spanish, Chinese (Mandarin), German, French, and Indian (Hindi). These languages were selected because Americans have strong folk artifacts for each of these specific accents (Lindemann 2005; Section 2.6). In addition to this, some of these languages are representative of broad artifacts that represent a variety of accents; for example, the Chinese accent also epitomizes an Asian artifact, and the Spanish accent symbolizes a South American artifact.

A total of 22 NNAE speakers participated. In order to narrow the speakers down to one male and one female per accent, three American linguists were consulted. Since phonetically trained native listeners have more sensitivity in discriminating voices than untrained listeners (Schiller and Koster 1998), it was assumed that American linguists would be able to recognize both when a speaker closely matched a folk artifact and when a speaker produced features that were typical of the accent. Precedent for this is found in e.g. Cunningham-Andersson 1996, who had a panel of dialectologists choose the most authentic speakers for listeners to hear. The linguists in the present research rated each speaker on a scale of 1-7 for how representative their voice was of the particular accent, where 1 was "Not at all representative" and 7 was "Very representative." The lowest cumulative score possible was 3; the highest possible score was 21. The higher a speaker's

score, the more representative they were judged to be of that artifact. The male and female with the highest cumulative score for each language were included in the study. Table 3 below provides individual overall scores of the speakers who were included. Scores of these ten speakers averaged 17.55, with a wide range of 12-21 indicating fairly broad variation among the representativeness of speakers.

Native Language	Speaker	Sex	Score
Chinese	F10	Female	18.5
	F04	Male	18.5
French	F21	Female	12
	F16	Male	21
German	F33	Female	16
	F19	Male	17
Indian	F07	Female	18
	F05	Male	18.5
Spanish	F30	Female	16.5
	F29	Male	19.5
Average			*17.55*

Table 3 Representativeness ratings for top male and female speakers of each language.

3.3.2 AE Speakers

The second group of speakers consisted of native speakers of American English (AE). They were recorded reading the same passage as the NNAE speakers, in three guises: in their "regular" voices, doing free (spontaneous) imitations, and doing modeled imitations (after hearing authentic accents). This is similar to the methodology of Torstensson et al., who compared speakers' imitations before and after training with authentic accents.

Twenty-six AE speakers were recruited from introductory linguistics courses and compensated with course credit for participating. They were recorded in a sound booth reading the same passage as the NNAE speakers. The AE speakers first read the passage three times in their regular voice and rated which reading they

thought was best. This was used as the speakers' "normal" voice. They then read the same passage while imitating the five NNAE accents (Chinese, French, German, Indian, and Spanish). Speakers were allowed to practice as much as they wanted. They were instructed to produce a natural and convincing accent, as opposed to a caricature, and informed that their goal was to convince future listeners that theirs was an authentic accent. These performances are the speakers' *free imitations*. In order to keep the imitations distinct, between each imitated passage the speakers "reset" to their regular voices by reading a short passage from a well-known folktale (*Goldilocks and the Three Bears*, see Appendix B).

After completing these five imitations, speakers performed the imitation task again. This time, however, they first heard authentic speakers with these accents. Recordings of the first group of participants (two NNAE speakers from each language) were presented, and the AE speakers performed their accent imitations after hearing each pair. These performances are the speakers' *modeled imitations.* The short reset passage between imitations was from another folktale (*The Little Boy Who Cried Wolf*, see Appendix B). Speakers concluded by answering questions about their familiarity with these accents.

In order to establish which speakers to use for each accent, the same group of three American linguists evaluated the speakers. The linguists rated how believable each imitation was, with 1 being "Not at all believable" and 7 being "Very believable." Each speaker was then judged twice for each accent: once for their free imitation and once for their modeled imitation. A speaker could have an overall free or modeled score of 3-21, and a cumulative total score of 6-42. The higher a

speaker's score, the more believable are their imitations. The results for the speakers who were included in the study are shown in Table 4 below. The mean total score was 26.55, with a range of 20.5-36.5. The wide range of these highest scores hints at the variety in imitation skill among speakers.

Target Accent	Speaker	Sex	Free Score	Modeled Score	Total Score
Chinese	A07	Female	15.5	13.5	29
	A12	Male	14.5	16	30.5
French	A19	Female	11	10.5	21.5
	A28	Male	18	18.5	36.5
German	A14	Female	10.5	10.5	21
	A23	Male	16.5	18	34.5
Indian	A18	Female	12	11	23
	A06	Male	13.5	14.5	28
Spanish	A15	Female	10.5	10	20.5
	A27	Male	11	10	21
Average			*13.3*	*13.25*	*26.55*

Table 4 Believability ratings for the final AE speakers.

3.3.3 Listeners

The third group of participants consisted of AE listeners. They were recruited in two ways: through a local university, where they received course credit for participation, and through social networking, where there was no compensation. The university was a different one from where the speakers were recruited, to ensure that listeners did not recognize individual speakers. Social networking included global Facebook groups and friend-of-a-friend email chains. This method was included because in the pilot study it had succeeded in quickly attracting a large amount of participants from a variety of locations.

This survey was conducted over the Internet and was therefore done at the listeners' convenience, whenever and wherever they had computer and Internet

access. A total of 100 listeners completed the survey. An additional 295 surveys were partially completed and discarded. Of the 100 complete surveys, 48 participants were bilingual or non-native speakers of American English (NNAE); these results were not included as the goal was to investigate AE perceptions and artifacts. The final listener group consisted of 52 participants who were native speakers of AE (13 from the class and 39 from the internet).

3.3.4 Procedure

Several methodological decisions were influenced by the pilot study, as described in section 2.6. I included multiple speakers for each language to avoid individual idiosyncrasies[11]. To facilitate quantitative evaluation, this survey employed forced-choice instead of free response questioning, a rating scale to measure familiarity, and a rating scale to measure confidence (referred to as "certainty"). This is similar to Derwing & Munro 1997, who also used forced choice identification and a rating scale for familiarity. For the present experiment, a forced choice task was acceptable because the choices were based on the artifacts found in the pilot study, which were consistent with Lindemann 2005. The accents listeners heard were those which the AE population has strong folk artifacts for, to assist them in identification and ensure that the labels provided by the forced choice questioning truly reflected listener categorization schemes.

[11] The trade-off in experiment time and number of speakers necessitated a limit of two authentic speakers and two imitated speakers per accent. Admittedly, even more speakers would be a great improvement.

This experiment had two conditions, in which listeners were randomly placed. In the first condition, listeners heard authentic accents and free imitations; in the second condition, listeners heard the same authentic accents and modeled imitations. By using the same speaker in different guises, this approximates the matched guise technique (e.g., Lambert et al. 1965, Lambert 1967), but unlike in matched guise the listener heard each speaker in only one guise. This was done so that listeners did not hear the same voice producing two different types of imitations, as it was determined that the two types of imitations were too similar for a traditional matched guise method. By comparing the results from the two conditions, we can address the fourth hypothesis: that listeners would react differently to free and modeled imitations.

The tasks posed to listeners in both conditions were similar to Neuhauser & Simpson 2007: identify the accent and determine its authenticity. First, listeners heard a voice and determined what the native language of that speaker was, choosing from a list of the five languages. The order of the voices was randomized. After each response, listeners rated their certainty regarding their decision. Figure 8 shows how this survey was presented to listeners. This accent identification task was done to address the second hypothesis: that listeners would be able to recognize the speaker's accent.

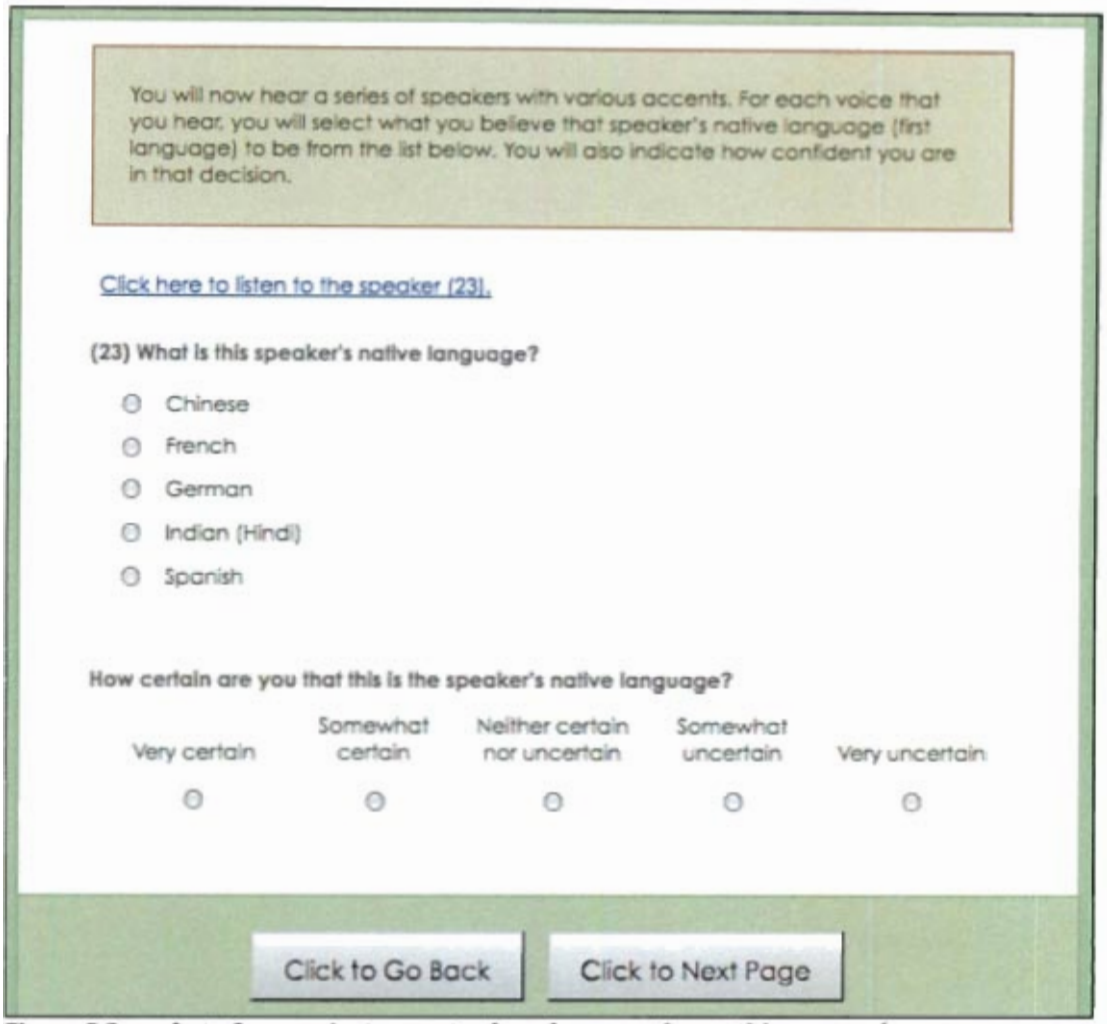


Figure 8 Snapshot of survey instrument, where language (accent) is assessed.

After hearing all of the voices, listeners were informed that some of the accents they heard were authentic (real) and some were imitated (fake). They then went through and made a binary decision for each voice as to whether it was a real or fake accent, and rated their certainty about each decision (see Figure 9). Voices were again randomized. This task addressed the third hypothesis: that listeners would not be able to differentiate between real and fake accents. Each survey concluded with listeners answering socio-demographic questions, in order to

determine if external factors such as familiarity were correlated with correct responses.

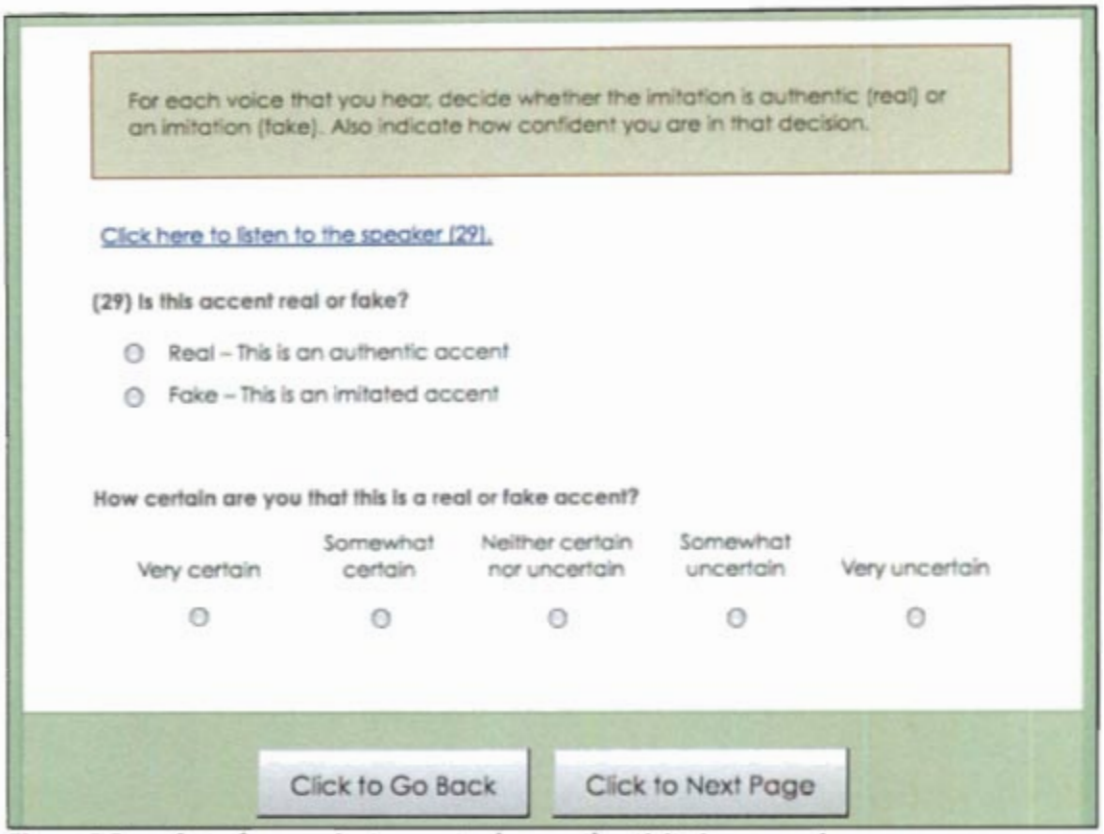


Figure 9 Snapshot of survey instrument, where authenticity is assessed.

3.3.5 Phonetic Analysis

These tasks produced two types of data to be analyzed: phonetic data from the speakers and survey responses from the listeners. The results of the speaker data are presented in Chapter 4, and the results of the listener data are presented in Chapter 5. This section describes how the phonetic data of speakers was collected.

Pitch and vowels were selected as the object of phonetic analysis. In a separate unpublished pilot study, I found that most speakers chose to vary their

pitch when asked to disguise their voice by imitating another accent; therefore, mean pitch was included as a feature of investigation here. The pitch range was included in order to determine if imitated pitch was exaggerated. Vowel formants and duration were assessed because it has been shown that they can be modified in imitations (e.g. Zetterholm 1997, Schilling-Estes 1998, Evans 2002), though the awareness of these modifications may vary. In particular, the availability of vowel features varies: while Zuengler 1988 finds that speakers are less "conscious" of vowel modifications than consonant modifications, Niedzielski and Preston 2000 note that "Vowels appear to mean more to the folk than consonants, at least in overt comment" (113). Speakers' awareness of vowel modifications presumably varies by target. It is necessary to again clarify the type of awareness that this study is investigating. The focus of this work is not on the mode of availability (high availability allows for overt comment), but on the mode of control (high control allows for modifications in imitation). Speakers with high control are more likely to modify their speech, and these modifications could play a role in listener perception, with varying modifications influencing listeners' identification of speakers.

In order to evaluate speakers' pitch, the recorded speech was entered into Praat (Boersma & Weenink 2008). Using the program's automated "Get pitch" queries, the mean, minimum, and maximum pitch from the sentence "I know all about what you've been doing" was extracted. The pitch range was calculated by subtracting the minimum pitch from the maximum pitch. This process was repeated for each speaker in each guise.

The vowel system of each speaker was also assessed, with the three AE guises (free, modeled, regular) kept separate. The vowels that were included are [i, ɪ, eɪ, ɛ, æ, aɪ, aʊ, ɑ, ɔ, oʊ, u][12], as shown in Table 5[13].

IPA	Symbol
i	iy
ɪ	I
eɪ	ey
ɛ	E
æ	ae
aɪ	ai
aʊ	aw
ɑ	a
ɔ	ou
oʊ	ow
u	uw

Table 5 Vowels included in analysis, with the alternate symbols which are used in this text.

Two Praat scripts were run to extract the vowel data for each speaker. One script measured the duration of the vowels, based on beginning and end points which had been marked by hand. The other script measured the F1, F2, and F3[14] at two points (25% and 75%) within each vowel. Auditory and visual comparison of results revealed that the glides were not very different across guises, and so only the nuclei (first measurements) were included in the statistical testing. For between-speaker tests that required normalized data, NORM (Thomas & Kendall 2007) was

[12] While all the vowels were measured, each vowel had to have a minimum of 3-4 quality tokens to be included in the analysis, and therefore [ʊ, ʌ, ɔɪ] were excluded from the analysis.

[13] The symbols in this table were used for the vowels that were analyzed, since some of the programs used for analysis were not IPA-compatible, and these symbols are used instead of IPA throughout the rest of this dissertation.

[14] The F3 measurement was only included in normalizing, and not in the analysis.

used to produce results with the Bark Difference method. This vowel-intrinsic normalization technique was implemented because it was unclear how much the vowel spaces might vary across accents, and the Bark Difference method is not skewed by differences between dialects with different vowels.

The statistical analyses in this research were performed using SPSS 16.0 with an alpha level of .05. The testing of these specific acoustic measurements (pitch mean and range, vowel duration and mean formants) is discussed in the following chapter.

3.4 Conclusion

This chapter presented the motivation, hypotheses, and methodology of the present research. It discussed how this study builds on two previous studies of imitation, the hypotheses regarding speaker and listener performance, and the details of the experiment. The experiment addresses the concept of a folk artifact in several ways. The linguists' judgments essentially rate similarity to folk artifacts; the speakers' free imitations are productions of folk artifacts; the listeners' identifications show how folk artifacts are recognized; and the success ratings reveal listeners' confidence in their judgments of folk artifacts.

The following two chapters present the results of this research. In Chapter 4, the phonetic details of the free and modeled imitations are compared to each other, to the speakers' regular voices, and to the authentic accents. Chapter 5 presents an analysis of listeners' identification of the accent and authenticity of these voices.

Finally, Chapter 6 summarizes the results and relates them to folk artifacts and linguistic awareness.

Chapter 4 Speaker Results

This chapter presents the results from the phonetic analyses of pitch, vowel duration, and vowel formants. By assessing a select set of the phonetic details of speech, we can address the first hypothesis: that speakers' modeled imitations will be more similar to the authentic accents than their free imitations. Data from the AE speakers is analyzed in 4.1, and vowel and accent type (free imitation, modeled imitation, regular voice) emerge as significant factors, while language is much less important. The NNAE speakers are analyzed in 4.2, and language is only found to be significant as an interaction. The results of both the AE and NNAE speakers are compared in 4.3, where we see that the imitations are more similar to authentic accents than the regular voices. The results of this chapter are summarized in 4.4. While there are significant differences between the imitations and the regular voice, with the imitations generally produced more like the authentic accents, there are not significant differences between the free and modeled imitations. Since neither imitation is more similar to the authentic accents than the other, the hypothesis is not supported.

4.1 AE speakers

Each AE speaker was recorded reading the passage in three accent types, or guises: free imitation, modeled imitation, and regular voice. The phonetic data that

was analyzed in this research was the pitch mean and range, and the duration and first and second formants of vowels. Various Analyses of Variance (ANOVAs) were used to determine the significance of factors. An alpha level of .05 was used for all tests.

Pitch was measured and analyzed for all ten speakers in three guises each. In order to determine whether there were significant differences based on the factors of *language* and *accent type,*[15] I ran a two-factor mixed ANOVA, A x (B), where

A = *Language*	Between-subjects	Chinese, French, German, Indian, Spanish
B = *Accent Type*	Within-subjects	Free, Modeled, Regular

This test was run separately for both the mean pitch and the mean pitch range.

Three aspects of speakers' vowels were investigated: duration, the first formant (F1), and the second formant (F2). This involved measuring 11 vowels and calculating the means for all ten speakers in each guise. In order to determine whether there were significant differences based on the factors of *language, accent type,* and *vowel,* I ran a three-factor mixed ANOVA, A x (B x C), where

A = *Language*	Between-subjects	Chinese, French, German, Indian, Spanish
B = *Accent Type*	Within-subjects	Free, Modeled, Regular
C = *Vowel*	Within-subjects	a, ae, ai, aw, E, ey, I, iy, ou, ow, uw

This test was run separately for the duration data, the F1 data, and the F2 data.

[15] Factors and their interactions are italicized.

4.1.1 Pitch

Results of the ANOVA for mean pitch are shown in Table 6 below. Pitch was significantly different based on *accent type* ($F(2, 10) = 8.418$, $p = .007$, $\eta^2 = .350$). *Language* and the interaction of *accent type x language* were not significant.

Source	df	F	P	η^2
Language	4	.075	.987	.057
Accent Type	2	8.418	.007	.350
Accent Type x Language	8	2.664	.074	.443

Table 6 Results from ANOVA of AE pitch, with factors of *language* and *accent type*. Significant *P* values are highlighted.

Post-hoc pairwise comparisons of *accent type* showed that the free imitation was significantly different from the regular guise ($p = .012$) (see Table 7). The differences between the two imitation types and between the modeled imitation and regular guise were not significant. Figure 10 depicts the differences in *accent type*. It shows that the free pitch was higher than the modeled pitch, which was higher than the regular pitch.

Accent Type	Mean Difference	Standard Error	P
Free ~ Modeled	9.199	4.359	.266
Free ~ Regular	18.095	3.579	.012
Modeled ~ Regular	8.896	5.154	.435

Table 7 Post-hoc pairwise comparison of *accent type* in pitch.

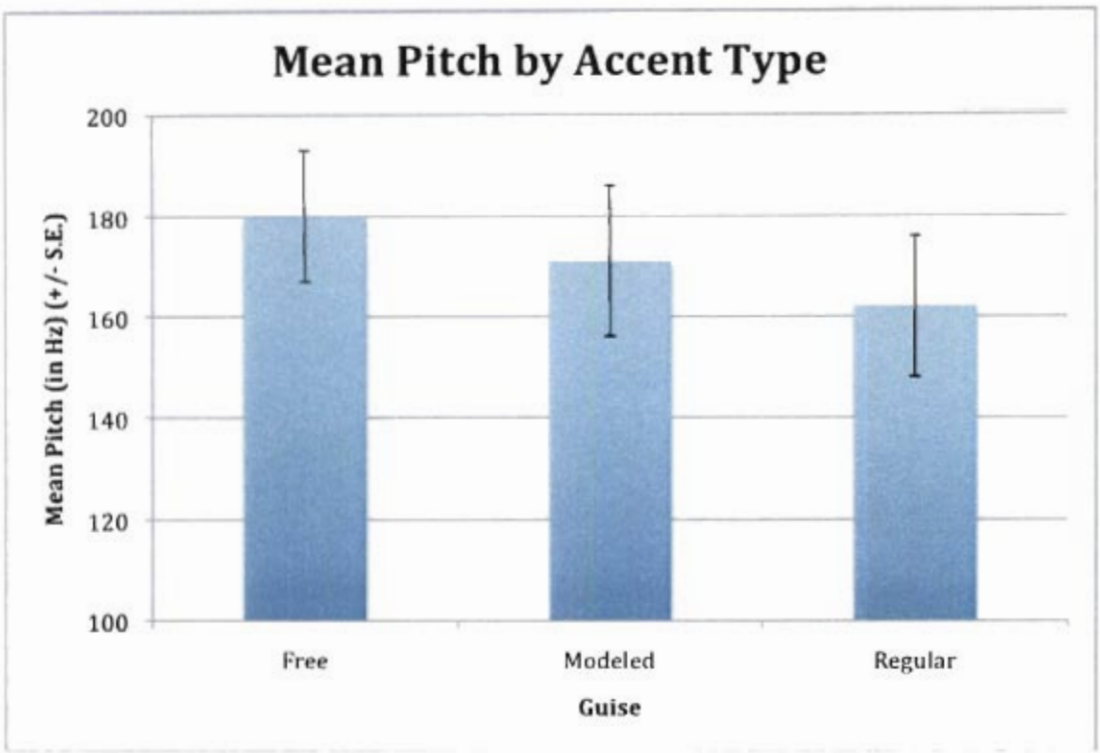


Figure 10 Means of AE pitch, by *accent type*.

An ANOVA for pitch range was also run. None of the results were significant, indicating that pitch range did not differ significantly by *accent type* or *language*.

4.1.2 Duration

Results of the ANOVA for vowel duration are shown in Table 8 below. Duration differs significantly by *vowel*, as expected (F (10, 50) = 38.077, p = .000, η^2 = .601). The other main effects of *accent type* and *language* were not significant, nor were any interactions including *language*. The interaction of *accent type x vowel* was significant (F (20, 100) = 2.674, p = .001, η^2 = .027).

Source	df	F	P	η^2
Language	4	3.310	.111	.729
Accent Type	2	.824	.466	.013
Accent Type x Language	8	.873	.568	.053
Vowel	10	38.077	.000	.601
Vowel x Language	40	.852	.698	.053
Accent Type x Vowel	20	2.674	.001	.027
Accent Type x Vowel x Language	80	1.193	.201	.047

Table 8 Results from ANOVA of AE duration, with factors of *language, accent type,* and *vowel.*

Figure 11 below depicts mean vowel duration by accent type. It is clear that for several vowels, the regular guise is longer than the imitations: this is the case for [a, ae, I, ow, ou, uw]. For the vowels [ai, E, ey, iy] the modeled imitation is longest. The remaining vowel was [aw], whose free imitation was shorter than the modeled imitation and regular voice.

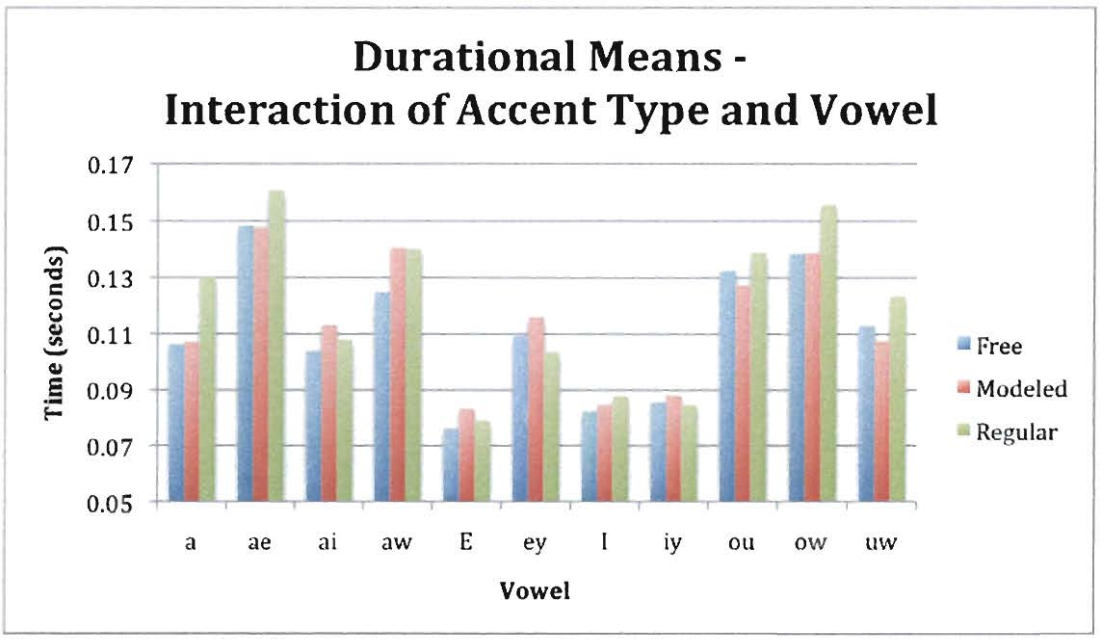


Figure 11 Means of AE duration, by *accent type* and *vowel.*

Figure 12 provides an alternative way to visualize these variations in mean duration by *accent type.* It shows the difference between the regular guise and the

imitations, where the difference is calculated as mean regular duration minus mean imitated duration, and the baseline of zero represents the regular guise. Bars that are below the baseline indicate how much shorter the imitated vowels are than the regular vowels; bars that are above the baseline indicate how much longer the imitated vowels are than the regular vowels.

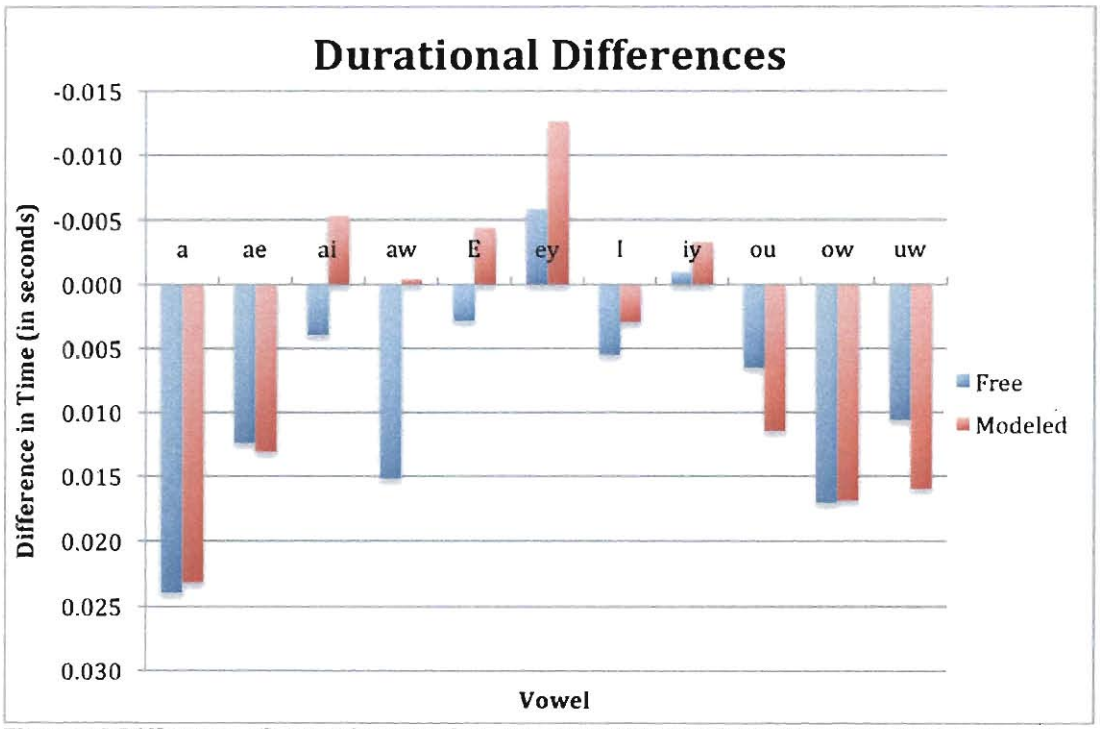


Figure 12 Difference of mean duration between imitations and regular voice. Bars represent the imitations and the baseline of zero represents the regular voice. Bars going downward indicate imitated vowels with a shorter duration than the regular voice; bars going upward indicate imitated vowels that had a longer duration than the regular voice.

In order to determine whether there were durational patterns based on the type of vowel, I ran another three-factor mixed ANOVA, replacing the *vowel* factor with *vowel type*: monophthong or diphthong (see Table 9). The main effects of *accent type* and *vowel type* were both significant (*accent type* F (2, 10) = 26.511, p = .000, η^2 = .343; *vowel type* F (1, 5) = 62.094, p = .001, η^2 = .429). The interaction of

accent type x vowel type was also significant (F (2, 10) = 5.804, p = .021, η^2 = .029). Again, *language* and interactions including *language* were not significant.

Source	df	F	P	η^2
Language	4	.841	.554	.007
Accent Type	2	26.511	.000	.343
Accent Type x Language	8	.866	.572	.057
Vowel Type	1	62.094	.001	.429
Vowel Type x Language	4	.920	.519	.029
Accent Type x Vowel Type	2	5.804	.021	.029
Accent Type x Vowel Type x Language	8	.414	.888	.000

Table 9 Results from ANOVA of AE duration, with factors of *language, accent type,* and *vowel type.*

Post-hoc pairwise comparisons of *accent type* showed that both of the imitations were significantly different from the regular guise (free p = .003; modeled p = .020) (see Table 10). Figure 4 highlights the differences in *accent type*. Duration was shorter in the imitations than in the regular guise. This shortening is more pronounced in the difference between the free and regular guises, but is also significant in the difference between the modeled and regular guises. The difference between the free and modeled imitations is not significant.

Accent Type	Mean Difference	Standard Error	P
Free ~ Modeled	.013	.005	.110
Free ~ Regular	.035	.005	.003
Modeled ~ Regular	.022	.005	.020

Table 10 Post-hoc pairwise comparison of *accent type* in duration.

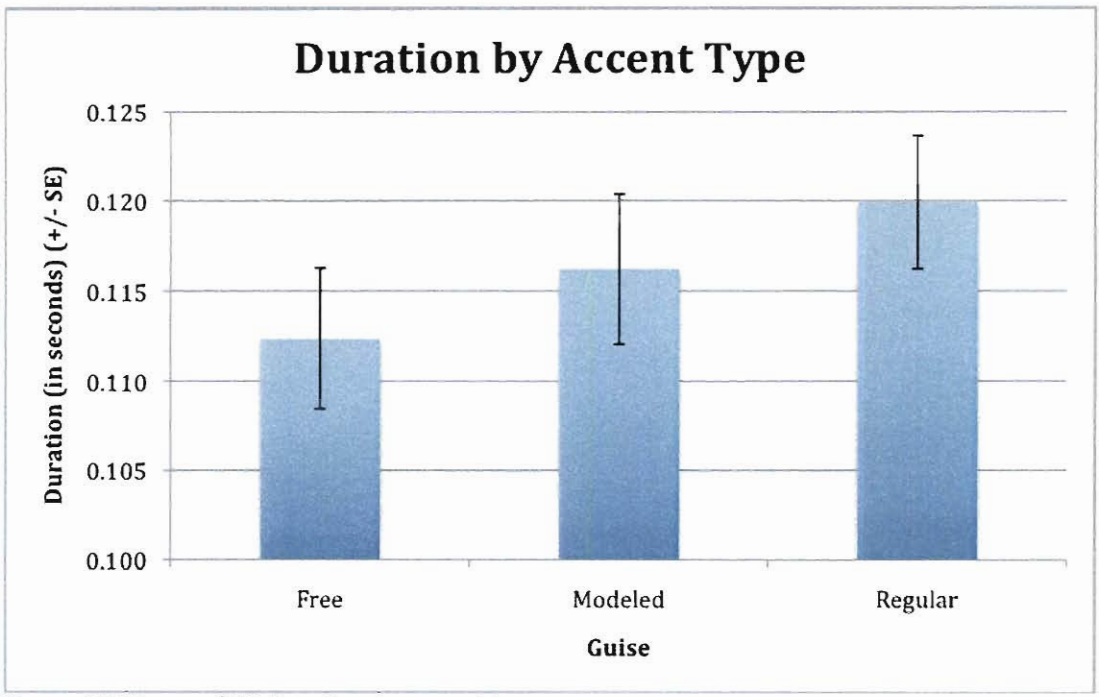


Figure 13 Means of AE duration, by *accent type.*

The difference between monophthongs and diphthongs is illustrated in Figure 14. As expected, monophthongs were significantly shorter than diphthongs.

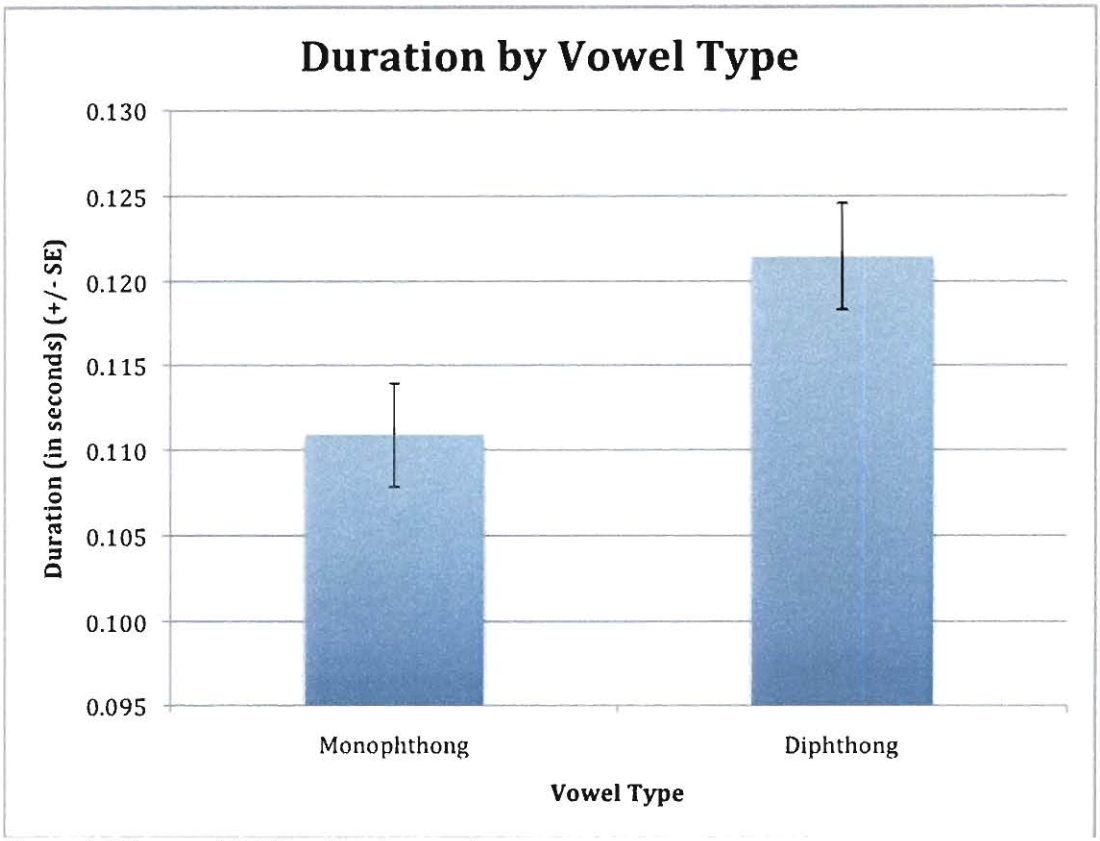


Figure 14 Means of AE duration, by *vowel type*.

Figure 15 presents the interaction of *accent type x vowel type*. It reveals that monophthongs were shortened more in the imitations than in the regular guise. This pattern is not the same with diphthongs, where the free diphthongs were shorter than the regular diphthongs, but the modeled diphthongs were slightly longer than the regular.

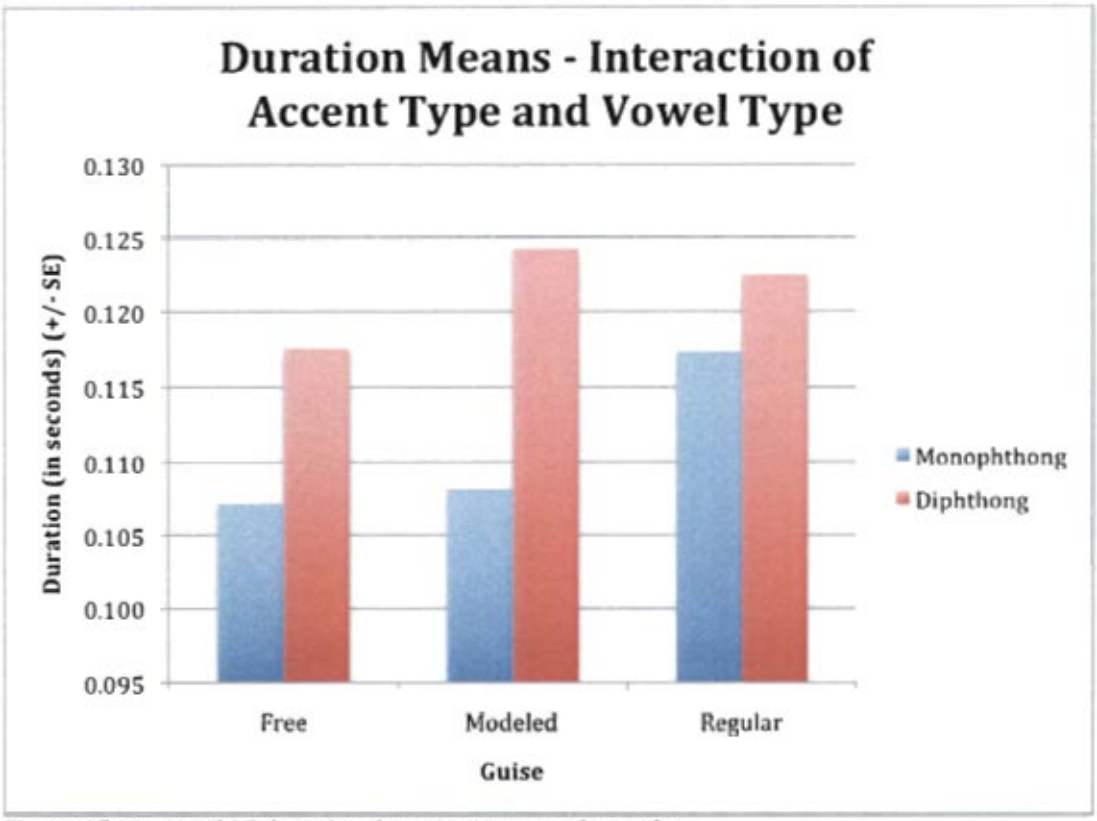


Figure 15 Means of AE duration, by *accent type* and *vowel type*.

To determine whether the difference between monophthong and diphthong length was significant, I ran a repeated-measures ANOVA. It showed a significant effect based on *accent type* (F (2, 18) = 3.816, p = .042, η^2 = .500; see Table 11). The difference in durations is illustrated in Figure 16.

Source	df	F	P	η^2
Accent Type	2	3.816	.042	.500

Table 11 Results of ANOVA comparing difference between monophthong and diphthong lengths.

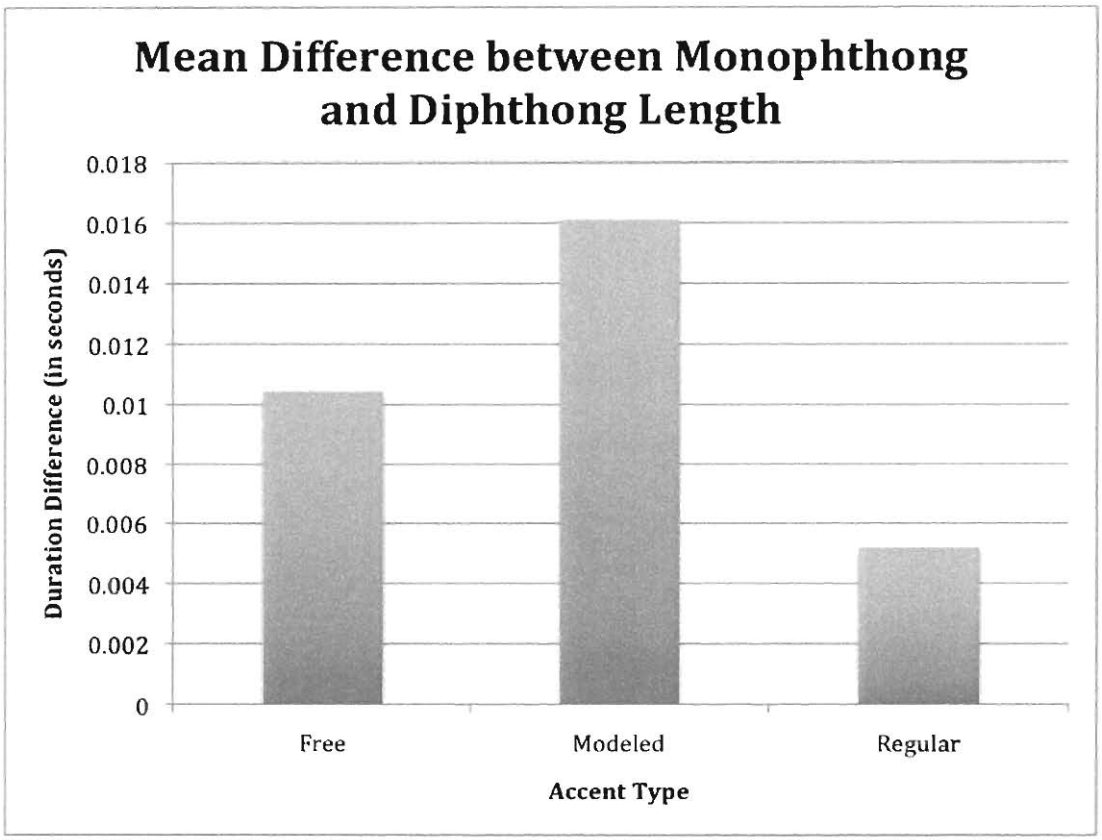


Figure 16 Mean durational difference in length between AE monophthongs and diphthongs.

To determine whether the difference in vowel duration was based on the speed at which speakers spoke, I measured the speech rate of each speaker in each guise (based on the length of a single sentence and calculated as syllables/second). I then ran a repeated-measures ANOVA on the speech rate. It showed a significant effect based on *accent type* (F (2, 18) = 4.189, p = .032, η^2 = .318; see Table 12).

Source	df	F	P	η^2
Accent Type	2	4.189	.032	.318

Table 12 Results of ANOVA comparing difference in speech rates.

The difference in durations is illustrated in Figure 17, where a higher speech rate denotes faster speech. The regular voices were faster than the free imitations,

which were faster than the modeled imitations. This is counter-intuitive, as one might imagine the regular voices would be the slowest since their vowels have the longest duration. Though the imitations have a slower speech rate, their vowels are shorter. Three separate Pearson correlations were performed on the variables of speech rate, vowel duration, and diphthong/monophthong difference, to see if any relationships existed; none of them were significant.

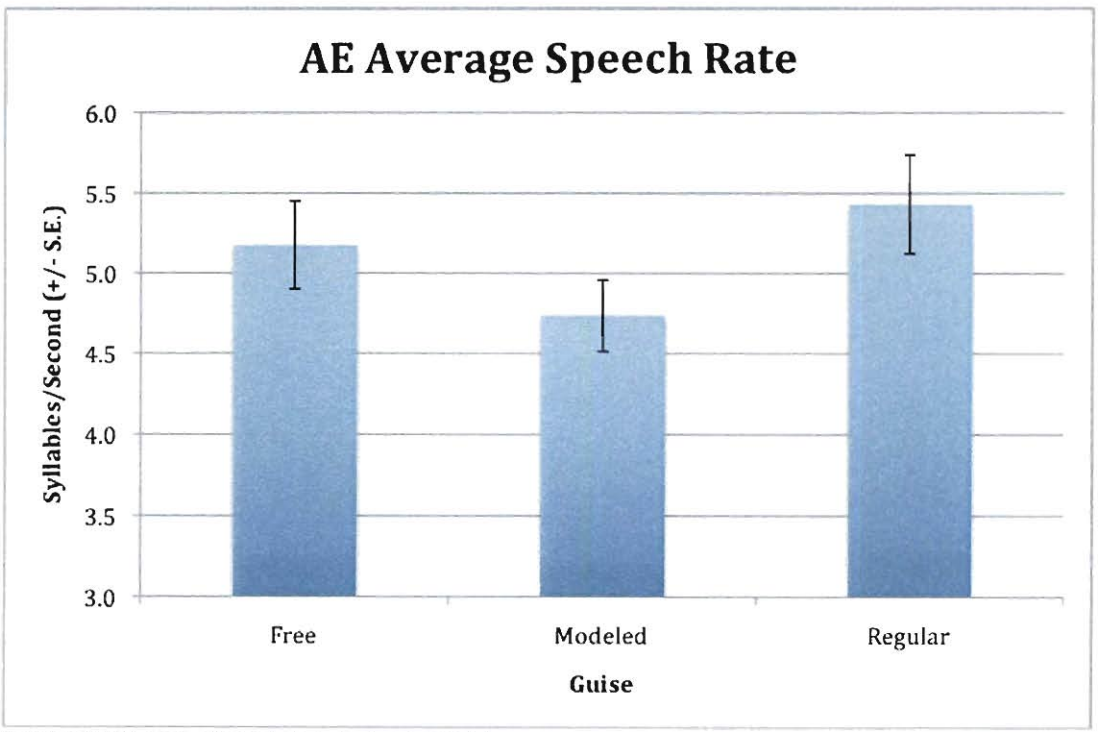


Figure 17 Means of speech rate, by *accent type*.

4.1.3 F1

The results for F1 can be seen in Table 13 below. There was a significant main effect for *accent type* (F (2, 10) = 32.13, p = .000, η^2 = .008); a significant main effect for *vowel* (F (10, 50) = 142.888, p = .000, η^2 = .873); a significant interaction of *accent type x language* (F (8, 10) = 5.870, p = .006, η^2 = .006); and a significant

interaction of *accent type x vowel* (F (20, 100) = 3.734, p = .000, η^2 = .005). *Language* as a main effect was not significant; nor were the interactions of *vowel x language* or *accent type x vowel x language.*

Source	df	F	P	η^2
Language	4	.190	.934	.001
Accent Type	2	32.138	.000	.008
Accent Type x Language	8	5.870	.006	.006
Vowel	10	142.888	.000	.873
Vowel x Language	40	.287	1.000	.007
Accent Type x Vowel	20	3.734	.000	.005
Accent Type x Vowel x Language	80	1.084	.349	.006

Table 13 Results from ANOVA for AE F1, with factors of *language, accent type,* and *vowel.*

Pairwise comparisons for the main effect of *accent type* are shown in Table 14 below. It is clear that the free imitation and the modeled imitation are both significantly different from the regular voice, but not from each other. This reveals that speakers are making significant F1 changes between their regular speaking voice and their imitations (p = .005 for both), but not between the two types of imitations.

Accent Types	Mean Difference	Standard Error	P
Free ~ Modeled	1.209	2.962	1.000
Free ~ Regular	-31.509	5.208	.005
Modeled ~ Regular	-32.718	5.325	.005

Table 14 Post-hoc pairwise comparison of *accent type* in F1.

The significant main effect of *accent type* is shown in Figure 18. The imitations both had an overall lower F1 than the regular voice. In the concept of a traditional F1 x F2 vowel space, this means that vowels are raised in imitations.

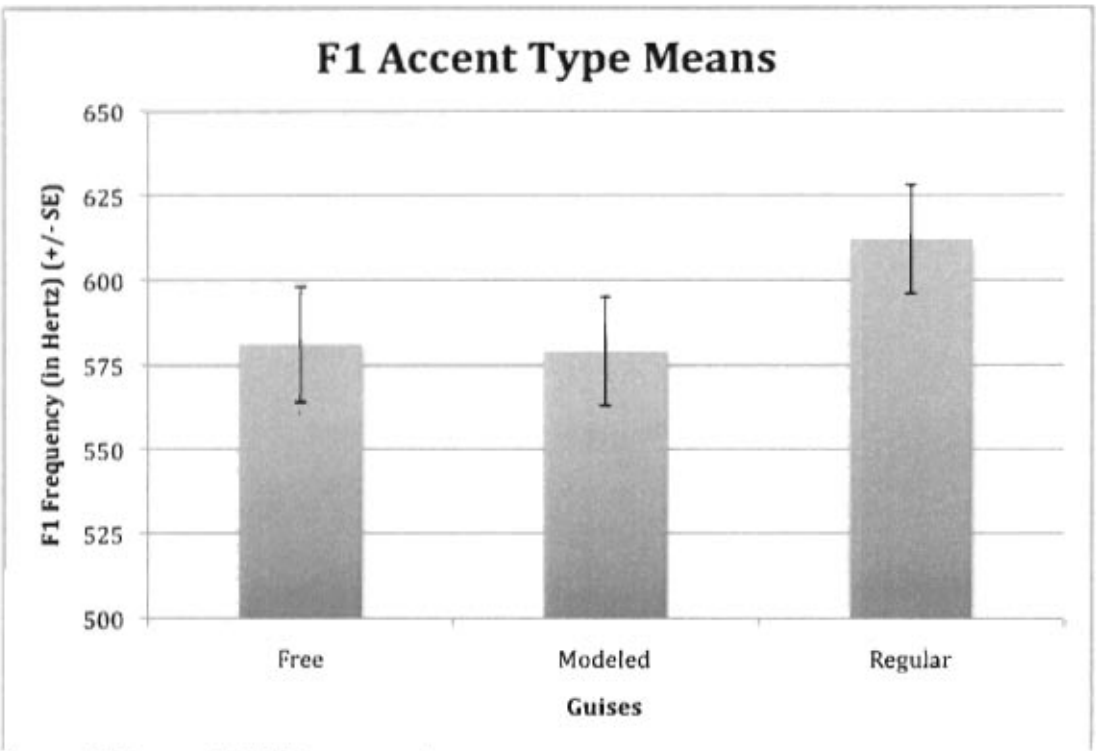


Figure 18 Means of AE F1 frequency, by *accent type*.

Figure 19 below sheds light on the interaction of *accent type x language*. F1 frequency is higher in the regular guise than in the imitations for German, French, and Indian. F1 is fairly steady across accent types for Chinese and Spanish.

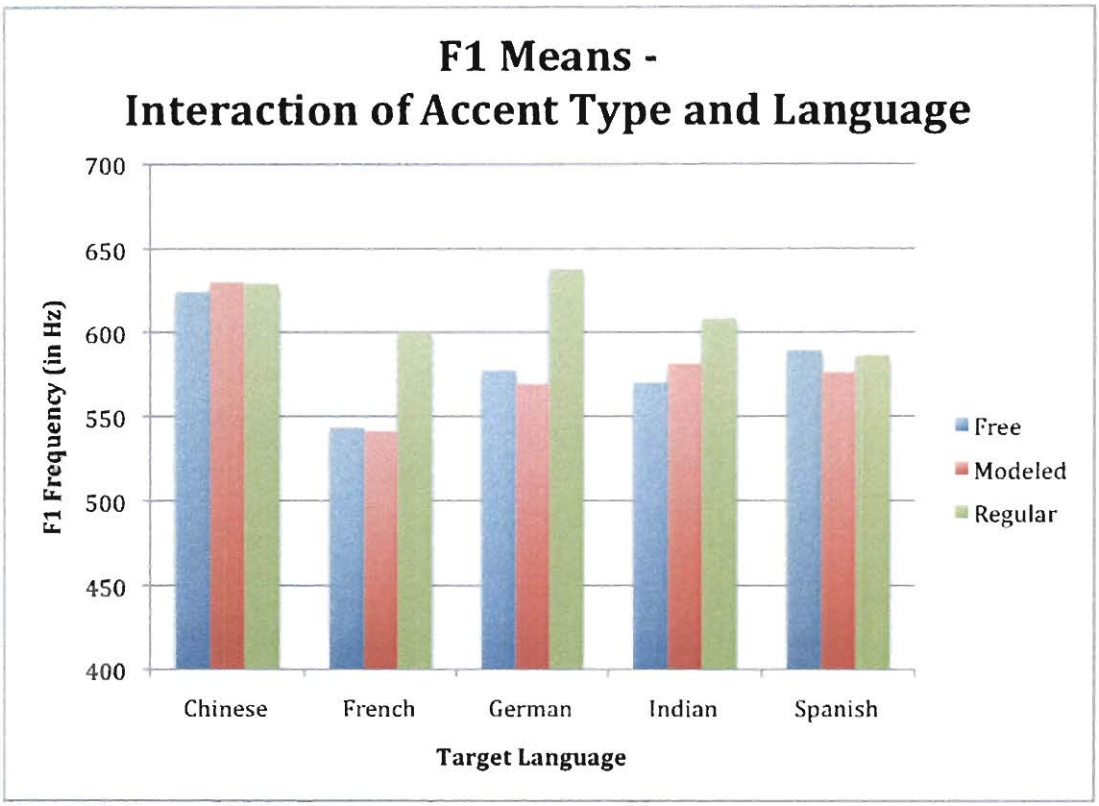


Figure 19 Means of AE F1 frequency, by *accent type* and *language*.

The interaction *of accent type x vowel* is shown in Figure 20 below. Most vowels have a slightly higher F1 in the regular guise: [a, E, ey, I, iy, ou, ow, uw]. There are some exceptions where F1 remains roughly the same across all three guises: the low vowels [ae], [ai], and [aw].

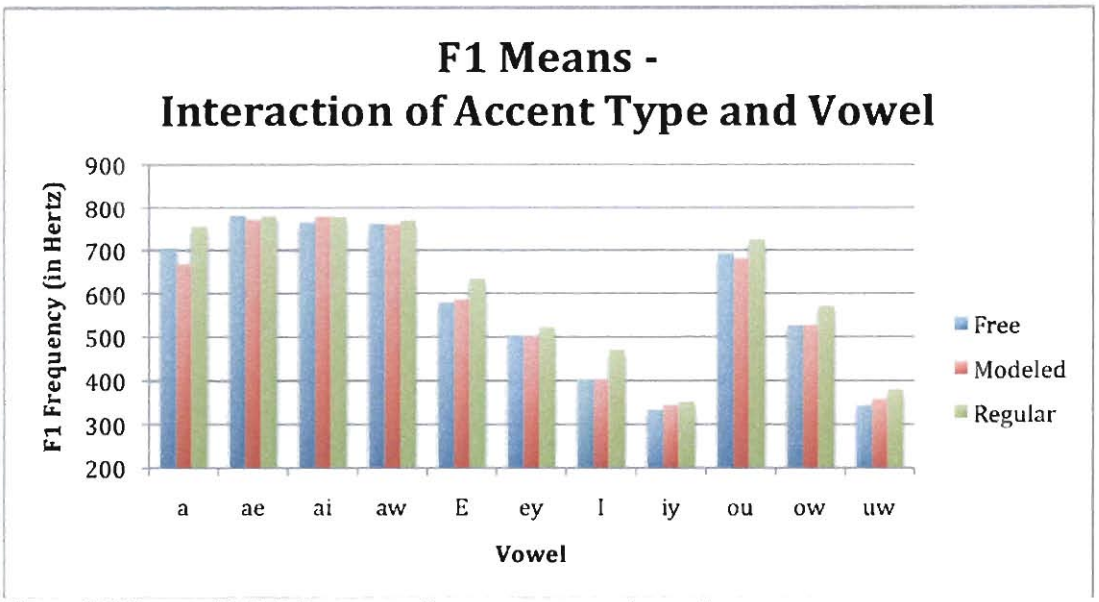


Figure 20 Means of AE F1 frequency, by *accent type* and *vowel*.

Figure 21 provides an alternative view of this difference by *accent type*. It shows the difference in mean F1 between the imitations and the regular voice, where the difference is calculated by subtracting mean imitation F1 from mean regular F1, and the regular F1 is used as a baseline for comparison. The higher the bar, the more the imitated vowel is raised in terms of a traditional F1 x F2 vowel space (and the lower the imitated frequencies in Hertz). Most imitated vowels had a raised F1.

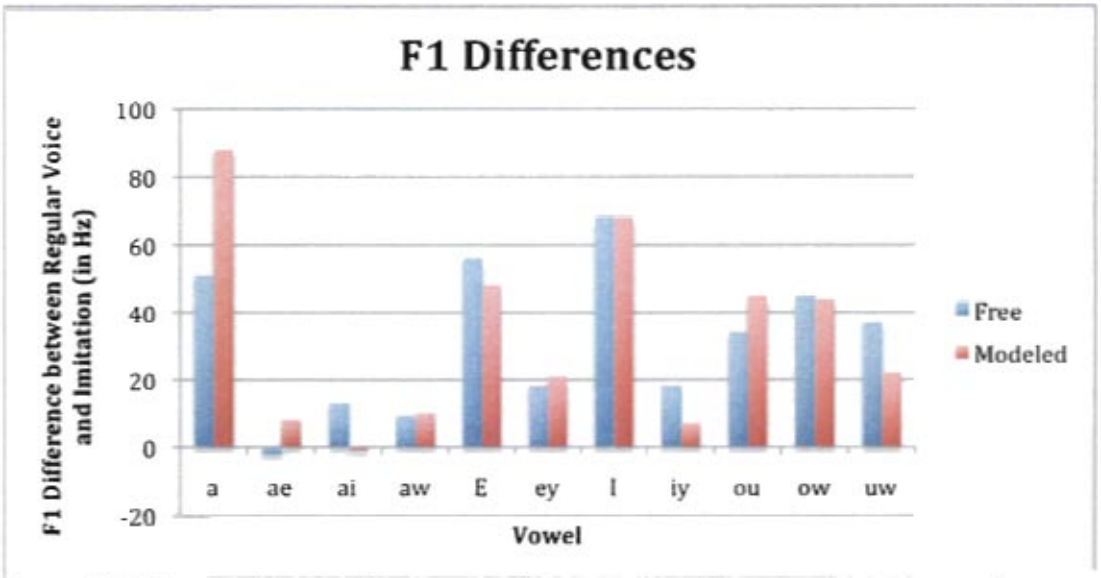


Figure 21 Difference of mean F1 between imitations and regular guise. Bars represent the imitations and the baseline of zero represents the regular voice. Bars going upward indicate the extent the vowels are raised in imitations from the regular voice.

4.1.4 F2

The results for F2 were similar to those for F1. *Accent type*, *vowel*, and the interaction of *accent type x vowel* were again significant (see Table 15). Specifically, there was a significant main effect for *accent type* (F (2, 10) = 10.663, p = .003, η^2 = .007); a significant main effect for *vowel* (F (10, 50) = 167.266, p = .000, η^2 = .871); and a significant interaction of *accent type x vowel* (F (20, 100) = 4.239, p = .000, η^2 = .019). *Language* and all interactions including it were not significant. Unlike F1, the interaction of *accent type x language* was not significant (p = .470).

Source	df	F	P	η^2
Language	4	.089	.982	.000
Accent Type	2	10.663	.003	.007
Accent Type x Language	8	1.036	.470	.003
Vowel	10	167.266	.000	.871
Vowel x Language	40	1.543	.073	.032
Accent Type x Vowel	20	4.239	.000	.019
Accent Type x Vowel x Language	80	.812	.833	.015

Table 15 Results from ANOVA for AE F2, with factors of *language, accent type,* and *vowel.*

Pairwise comparisons for *accent type* are shown in Table 16 below. Similar to F1, the free imitation is significantly different from the regular guise ($p = .017$) and the modeled imitation is significantly different from the regular guise ($p = .039$), while the free and modeled imitations are not significantly different from each other. As Figure 22 illustrates, the imitations both had a lower F2 than the regular guise.

Accent Types	Mean Difference	Standard Error	P
Free ~ Modeled	24.764	19.390	.773
Free ~ Regular	-57.600	12.449	.017
Modeled ~ Regular	-82.364	21.770	.039

Table 16 Post-hoc pairwise comparison of *accent type* in F2.

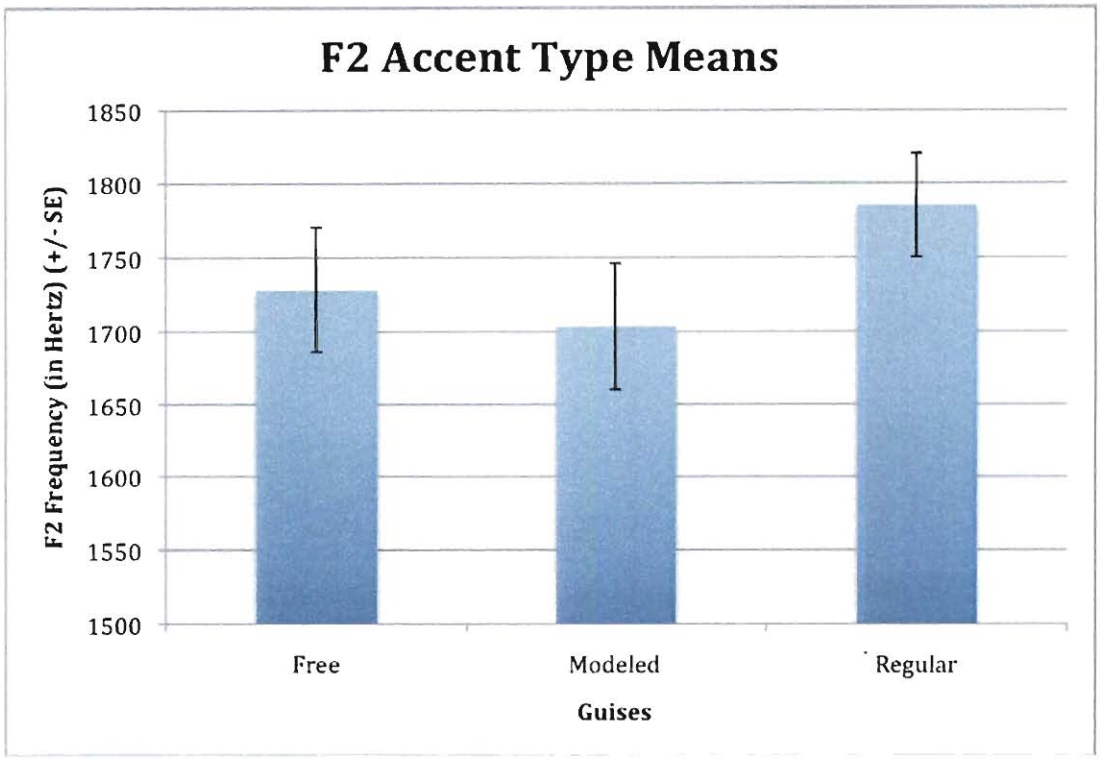


Figure 22 Means of AE F2 frequency, by *accent type*.

Figure 23 plots the interaction of *accent type x vowel*. F2 frequency is often higher in the regular guise. This is the case for most of the non-front vowels: [a, aw, ou, ow, uw]. The opposite is the case for the front lax [I] and [E], where F2 is lower for regular. For three vowels, the difference across guises appears negligible: the front tense [iy] and [ey], and [ai]. In the concept of a traditional F1 x F2 vowel space, this means that when vowels are modified in imitations, the back vowels are backing and some of the front vowels are fronting. [ae] is the exception to this trend, as it's getting backer instead of fronter. At first glance, it seems that the vowel space is expanding; however, that would be too simplistic of a conclusion. The fronting of front vowels is mostly due to [I] (and [E] to a lesser extent). While the back vowels are certainly backing, we cannot be sure whether this is in an effort to enlarge the

overall vowel space or simply to approximate the target or folk artifact, especially since the back vowels were fairly fronted in the regular voice and this backing movement returns them to more canonical positions.

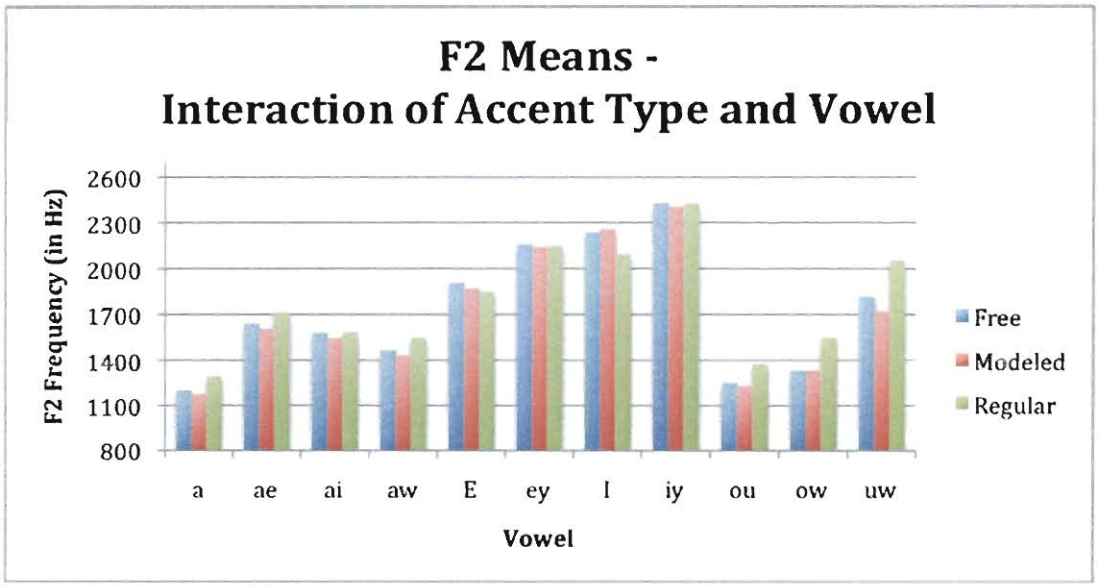


Figure 23 Means of AE F2 frequency, by *accent type* and *vowel*.

An alternative way to visualize these differences by *accent type* is presented in Figure 24 below. This shows the difference between imitations and the regular guise, where the mean imitated F2 is subtracted from the mean regular F2. The y-axis (zero) represents the regular guise. Where the bar is to the left of the axis, the imitated vowels are fronter than the regular vowels (in terms of a traditional F1 x F2 vowel space); where the bar is to the left of the axis, the imitated vowels are backer.

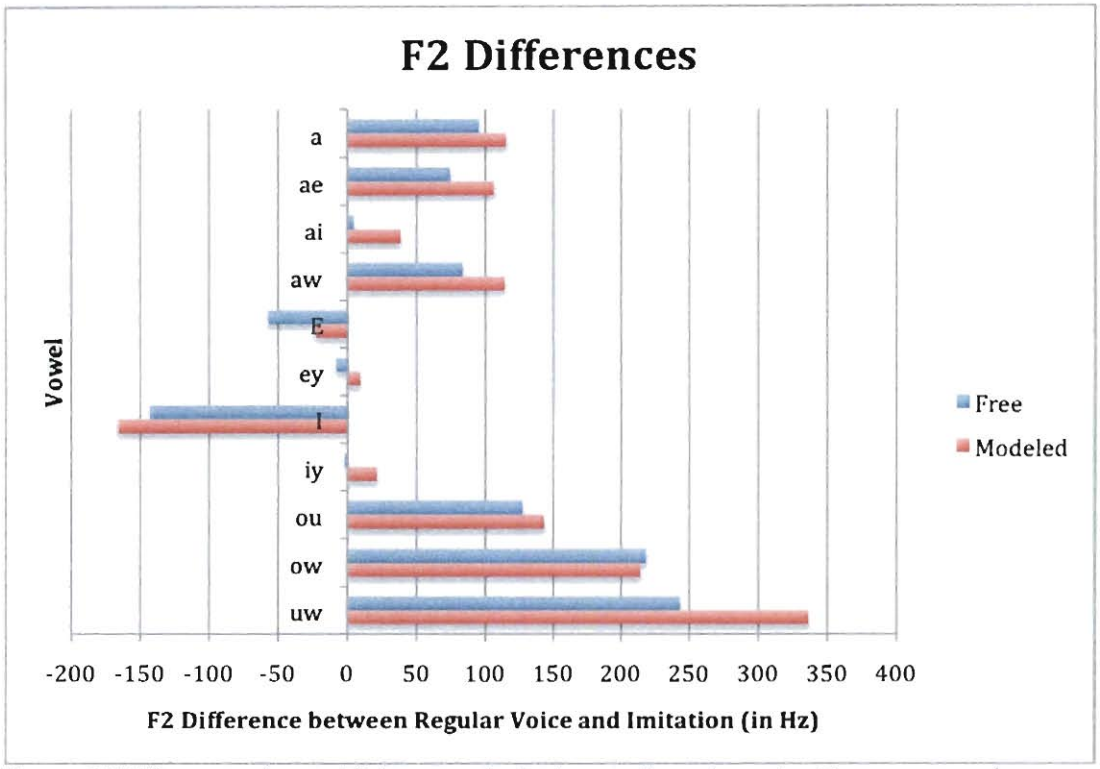


Figure 24 Difference of mean F2 between imitations and regular voice. Bars represent the imitations and the baseline of zero represents the regular voice. Bars to the left of the y-axis indicate imitated vowels that are fronted in comparison to the regular voice; bars to the right of the y-axis indicate imitated vowels that are backed in comparison to the regular voice.

While the interaction of *accent type x language* was not significant, the plot of it is still revealing (see Figure 25). We can see that for all languages except for French, F2 was highest for the regular voice. In addition to this, the modeled imitation F2 was always the lowest.

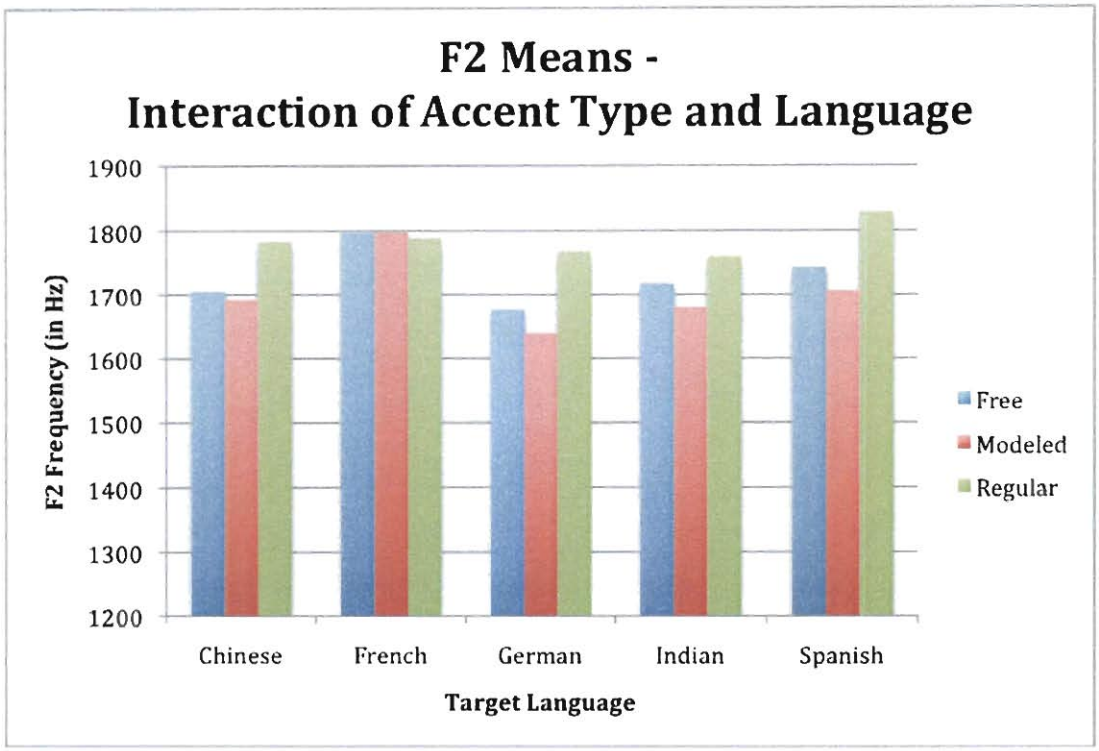


Figure 25 Means of AE F2 frequency, by *accent type* and *language*.

The ANOVA was repeated, replacing the factor of individual *vowel* with *vowel frontness* (front, mid, back) to see if that might influence the factors' significance (see Table 17). Results were similar to the original test, with significant factors of *accent type* (F (2, 10) = 11.503, p = .003, η^2 = .015) and *vowel frontness* (F (2, 10) = 523.56, p = .000, η^2 = .896), and an interaction between the two (F (4, 20) = 11.742, p = .006, η^2 = .023). In addition to this there was a significant interaction of *vowel frontness x language* (F (8, 10) = 4.113, p = .020, η^2 = .028).

Source	df	F	P	η^2
Language	4	.122	.968	.089
Accent Type	2	11.503	.003	.015
Accent Type x Language	8	1.045	.464	.005
Vowel Frontness	2	523.560	.000	.896
Vowel Frontness x Language	8	4.113	.020	.028
Accent Type x Vowel Frontness	4	11.742	.006*	.023
Accent Type x Vowel Frontness x Language	16	.948	.516*	.007

Table 17 Results from ANOVA for AE F2, with factors of *language, accent type,* and *vowel frontness.*
* Mauchley's Test of Sphericity showed significance, so I used more conservative Greenhouse-Geisser *p* values.

Pairwise comparisons revealed that the imitations were again both significantly different from the regular voice, but not from each other (see Table 18). A pairwise comparison for *vowel frontness* showed that the three groups (front, mid, and back) were all significantly different from each other, as expected (see Table 19).

Accent Types	Mean Difference	Standard Error	P
Free ~ Modeled	25.603	22.645	.929
Free ~ Regular	69.125	12.735	.009
Modeled ~ Regular	94.728	24.031	.033

Table 18 Post-hoc pairwise comparison of *accent type* in F2.

Vowel Frontness	Mean Difference	Standard Error	P
Front ~ Mid	677.825	27.717	.000
Front ~ Back	641.136	28.205	.000
Mid ~ Back	36.689	10.075	.045

Table 19 Post-hoc pairwise comparison of *vowel frontness* in F2.

Figure 26 shows the interaction of *vowel frontness x accent type*. Front vowels are very slightly higher in the imitations than in the regular voice, though the difference is minimal. Mid and back vowels are clearly lower in the imitations than

in the regular voice. This means that in a traditional F1 x F2 vowel space, the imitated back and mid vowels are backing.

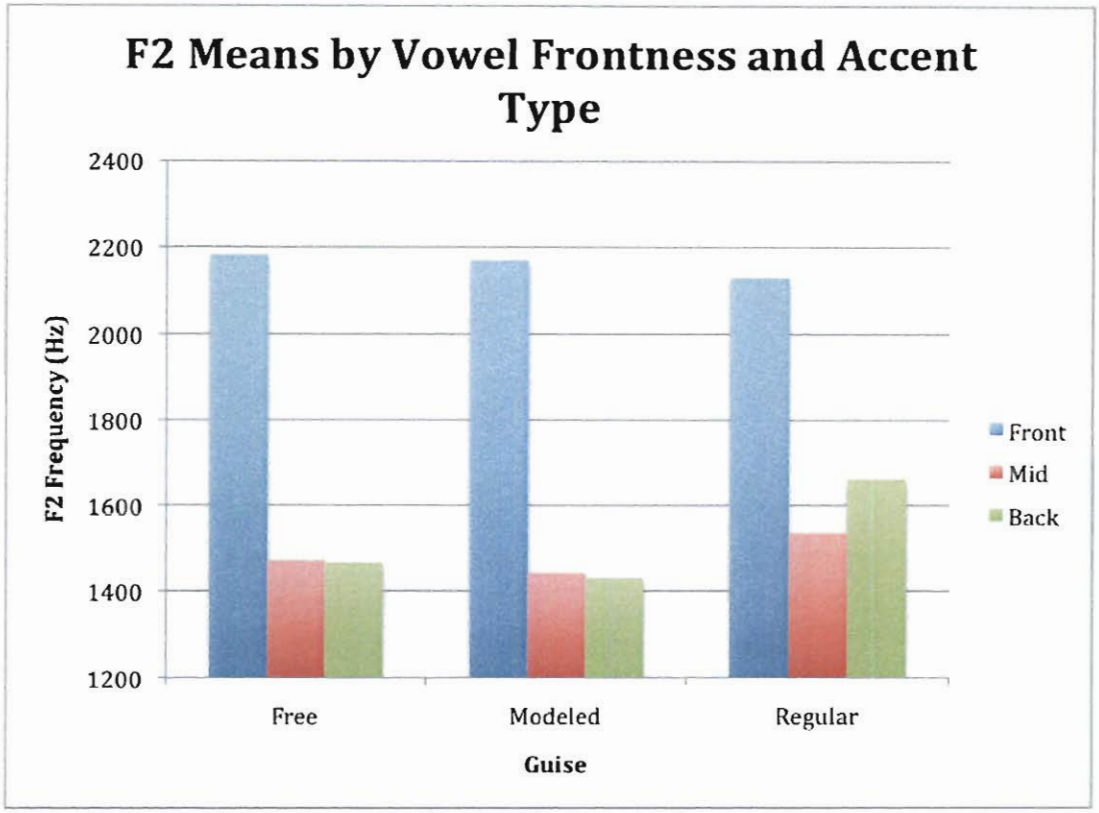


Figure 26 Means of AE F2 frequency, by *vowel frontness* and *accent type*.

Again, an alternative way to visualize these differences in presented in Figure 27, which shows the difference by both accent type and vowel type. This clarifies that, in terms of a traditional vowel space, the back vowels are backing quite a bit, the mid vowels are backing to a lesser extent, and the front vowels are fronting.

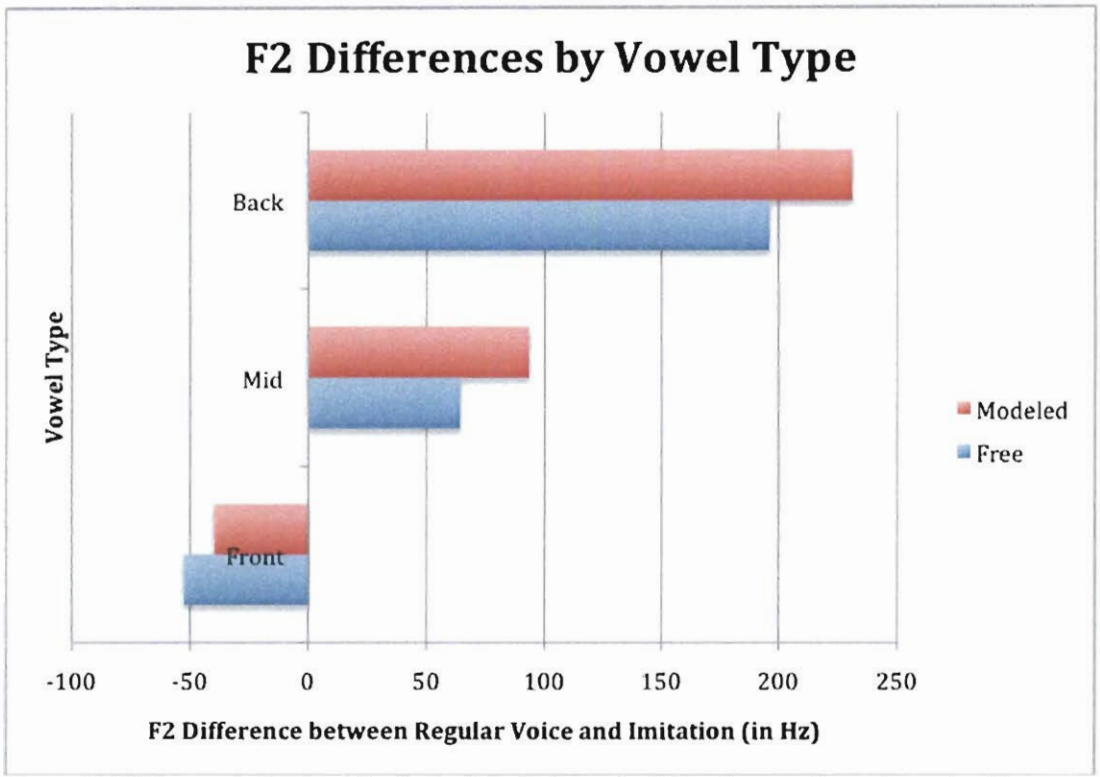


Figure 27 Difference of mean F2 of vowel types between imitations and regular voice. Bars represent the imitations and the baseline of zero represents the regular voice. Bars to the left of the y-axis indicate imitated vowels that are fronted in comparison to the regular voice; bars to the right of the y-axis indicate imitated vowels that are backed in comparison to the regular voice.

The interaction of *vowel frontness x language* is pictured in Figure 28. The variation in the back vowels is the most noticeable difference across languages.

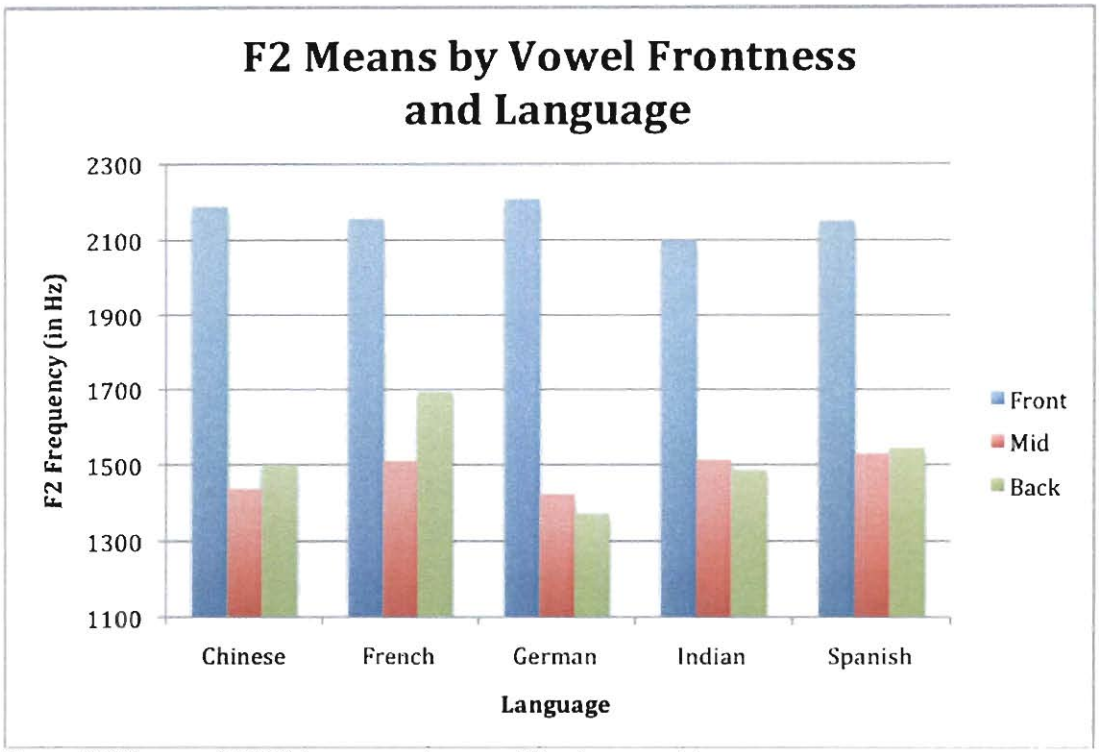


Figure 28 Means of AE F2 frequency, by *vowel frontness* and *language*.

4.1.5 Vowel Space

The following vowel plot, Figure 29, illustrates the differences between accent types for all the vowels. It combines the variations in F1 and F2 described above into a normalized plot. There is a noticeable difference in the F2 of the high back vowels [uw] and [ow], where the imitations are considerably backed in comparison to the regular vowels.[16] The front tense vowels [iy] and [ey] show little difference across accent types, with the imitations slightly higher than the regular. The non-low front lax vowels [I] and [E] reveal imitations that are consistently higher than the regular voice. The imitated [I] is also fronter. Moving down the

[16] Note that this type of normalization produces a chart which is slightly skewed from the traditional F1 x F2 space, representing e.g. [uw] as fronter than it would be in the unnormalized space.

vowel space, we see the regular [ae] to be slightly ahead of the imitated [ae]. The imitated [aw] is backed in comparison to the regular [aw]. [ai] does not seem to change consistently across accent types. Finally, both the imitated [a] and [ou] are higher and backer than the regular [a] and [ou].

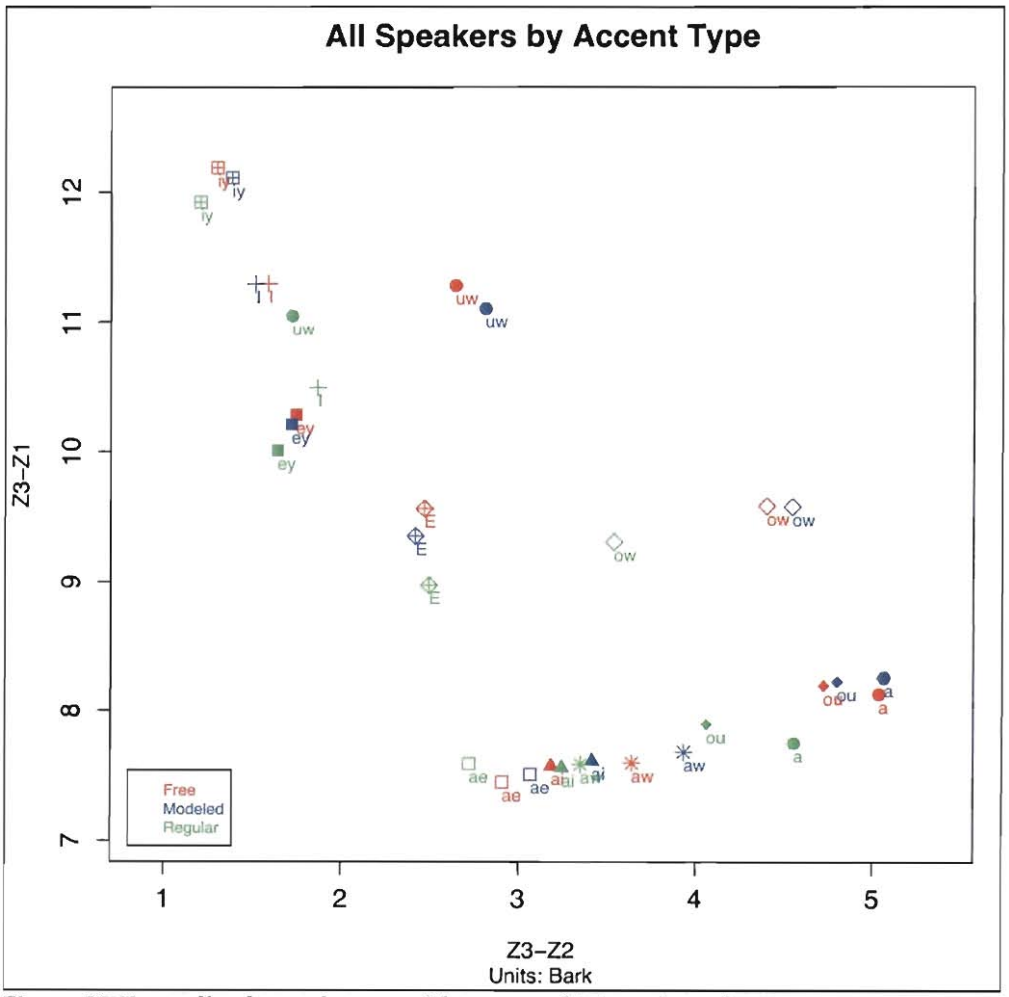


Figure 29 Normalized vowel space with means of AE speakers, by *accent type*.

4.1.6 Conclusions

Mixed ANOVAs were performed in order to determine whether there were significant differences among the AE speakers' guises for pitch and vowels.

Mean pitch varied significantly by *accent type* (free, modeled, regular). The imitations had a higher mean pitch than the regular voice, with the free imitation significantly higher than the regular. *Language* was not significant, not was the interaction of *accent type x language.* Pitch range did not vary significantly.

Duration varied significantly between individual vowels. The interaction of *accent type x vowel* was also significant. When analyzing by *vowel type* (monophthongs or diphthong) instead of by individual vowels, the main factors of *vowel type* and *accent type*, as well as the interaction of these two factors, were all significant. The monophthongs were shorter than diphthongs. Vowel duration of both imitations was significantly shorter than the regular voice. The interaction of *vowel type x accent type* revealed that the difference between monophthongs and diphthongs was much larger for the imitations than for the regular voice. Monophthongs in particular were shorter in the imitations. Speech rate, while it varied significantly across *accent types*, was not correlated with vowel duration or the difference between monophthongs and diphthongs. *Language* was never significant, neither as a main effect nor as an interaction, leading us to conclude that duration did not differ significantly across the five languages.

There were several significant effects and interactions for F1: *accent type, vowel, accent type x vowel,* and *accent type x language.* Pairwise comparisons showed that, within *accent types*, both the free imitation and modeled imitation were significantly different from the regular guise, though they were not significantly different from each other. These imitations both had a lower F1 than the regular guise. The *accent type x language* interaction revealed that the regular

guise had a higher F1 than the imitations for three languages in particular: German, French, and Indian. F1 was approximately the same across accent types for Spanish and Chinese. The *accent type x vowel* interaction showed that, with the exception of a few low vowels, the imitated vowels had a lower F1 than the regular vowels.

The results for F2 were similar. *Accent type, vowel,* and the interaction of *accent type x vowel* were significant. Repeating the ANOVA for *vowel frontness* instead of individual *vowel* found the same significant effects, with the addition of a significant *vowel frontness x language* interaction. A pairwise comparison again showed that free and modeled imitations were significantly different from the regular guise, but not from each other. These imitations both had significantly lower F2s than the regular voice. The *accent type x vowel (frontness)* interaction revealed that the imitated back and mid vowels were backed.

Combining F1 and F2 into a normalized vowel space for all speakers provided an illustration of the *accent type* differences. The imitated front vowels were higher than the regular, and the imitated back and low vowels were further back than the regular. The imitation vowels were consistently different from the regular guise vowels, though the free and modeled imitation vowels were not consistently different from each other.

4.2 NNAE speakers

The immediate targets for the modeled imitations were provided by authentic NNAE speakers. The phonetic data from these speakers was subsequently

evaluated. Separate ANOVAs were performed for pitch, duration, and combined F1 and F2.

4.2.1 Pitch

I ran a one-way repeated-measures ANOVA to determine if mean pitch varied by *language*. The results were not significant. There was also no significant difference by *language* in a one-way repeated-measures ANOVA of pitch range.

4.2.2 Duration

I ran a mixed A x (B) ANOVA, where A is *language* (between-subjects variable) and B is *vowel* (within-subjects). Results showed that there were significant differences by *vowel* (F (10, 50) = 26.523, p = .000, η^2 = .682), and that the interaction of *vowel x language* was also significant (F (40, 50) = 1.849, p = .020, η^2 = .193). The main effect of *language* was not significant (see Table 20).

Source	df	F	P	η^2
Language	4	.214	.920	.001
Vowel	10	26.523	.000	.682
Vowel x Language	40	1.849	.020	.193

Table 20 Results from ANOVA of NNAE duration, with factors of *language* and *vowel*.

To investigate if there were similar trends to the AE speakers, the test was repeated changing the *vowel* factor to *vowel type* (monophthong/diphthong). Once again, the main effect of *vowel type* and the interaction of *vowel x language* were significant, while the main effect of *language* was not (see Table 21).

Source	df	F	P	η^2
Language	4	.200	.928	.000
Vowel Type	1	52.719	.001	.500
Vowel Type x Language	4	15.036	.005	.500

Table 21 Results from ANOVA for NNAE duration, with factors of *language* and *vowel type*.

Figure 30 shows the durational differences of *vowel type*. Diphthongs are longer than monophthongs.

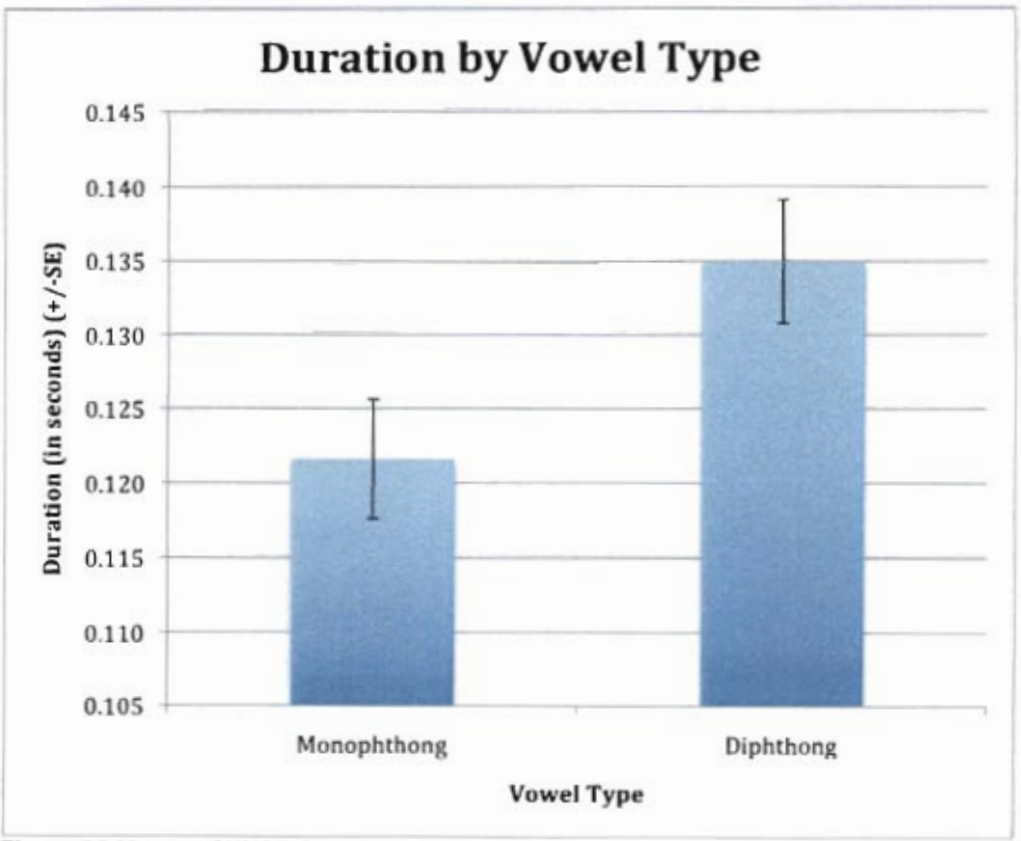


Figure 30 Means of NNAE duration, by *vowel type*.

Figure 31 highlights the interaction of *vowel type x language*. Chinese speakers showed almost no difference in the duration of monophthongs and diphthongs. The Indian and Spanish speakers' durational differences were slight. The French speakers had a notable difference between monophthongs and

diphthongs. Finally, the German speakers had a very large difference between monophthongs and diphthongs; the diphthongs were 1.4 times longer than the monophthongs.

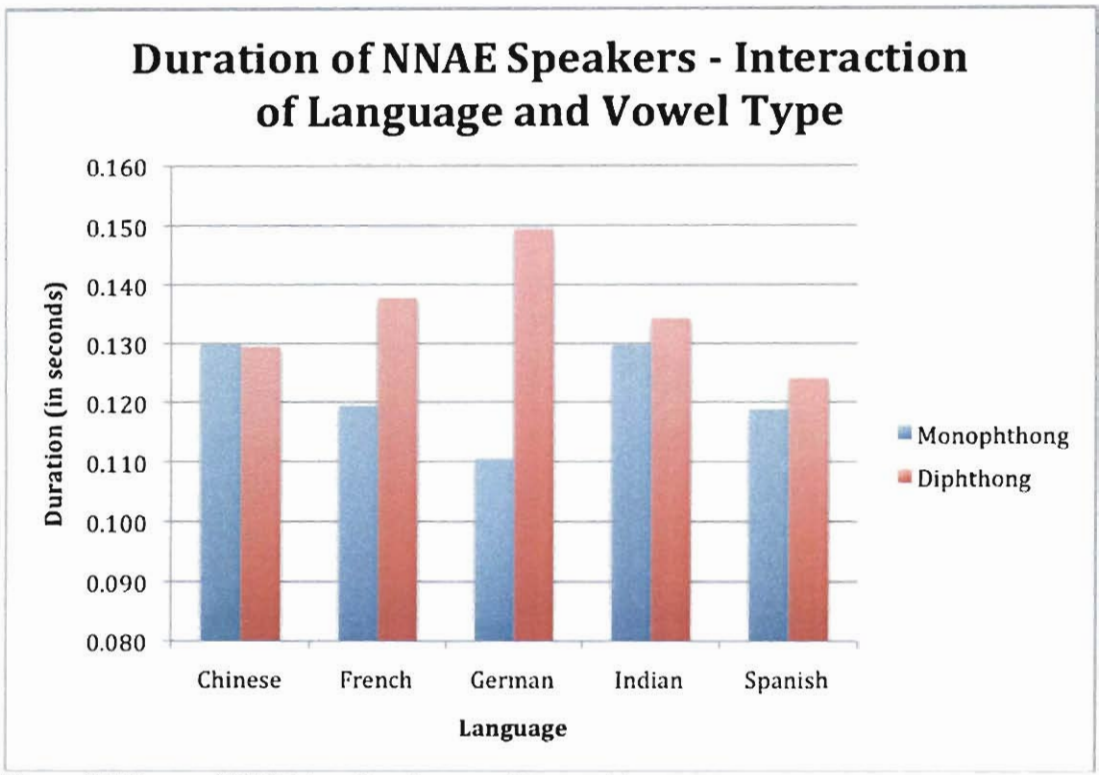


Figure 31 Means of NNAE duration, by *vowel type* and *language*.

Speech rate was also assessed. There were no significant correlations between speech rate and vowel length, or between either of those two features and monophthong/diphthong difference.

4.2.3 Vowel Space

In order to analyze the formants of NNAE speakers, I ran a mixed A x (B x C) ANOVA, where

A = *Language*	Between-subjects	Chinese, French, German, Indian, Spanish
B = *Vowel*	Within-subjects	a, ae, ai, aw, E, ey, I, iy, ou, ow, uw
C = *Formant*	Within-subjects	F1, F2

Results are shown in Table 22 below. As expected, the main factors of *vowel* and *formant* were significantly different (*vowel* F (10, 50) = 432.868, p = .000, η^2 = .838; *formant* F (1, 5) = 16.172, p = .010, η^2 = .001). The interaction of *vowel x formant* was also significant (F (10, 50) = 193.392, p = .000, η^2 = .136). The main factor of *language* was not significant, nor were the two-way interactions including it. The three-way interaction of *vowel x formant x language* was significant (F (40, 50) = 1.701, p = .038, η^2 = .005). Figure 32 depicts this interaction.

Source	df	F	P	η^2
Language	4	.338	.842	.002
Vowel	10	432.868	.000	.838
Vowel x Language	40	.721	.856	.006
Formant	1	16.172	.010	.001
Formant x Language	4	2.700	.153	.001
Vowel x Formant	10	193.392	.000	.136
Vowel x Formant x Language	40	1.701	.038	.005

Table 22 Results from ANOVA for NNAE F1 and F2, with factors of *language, vowel,* and *formant.*

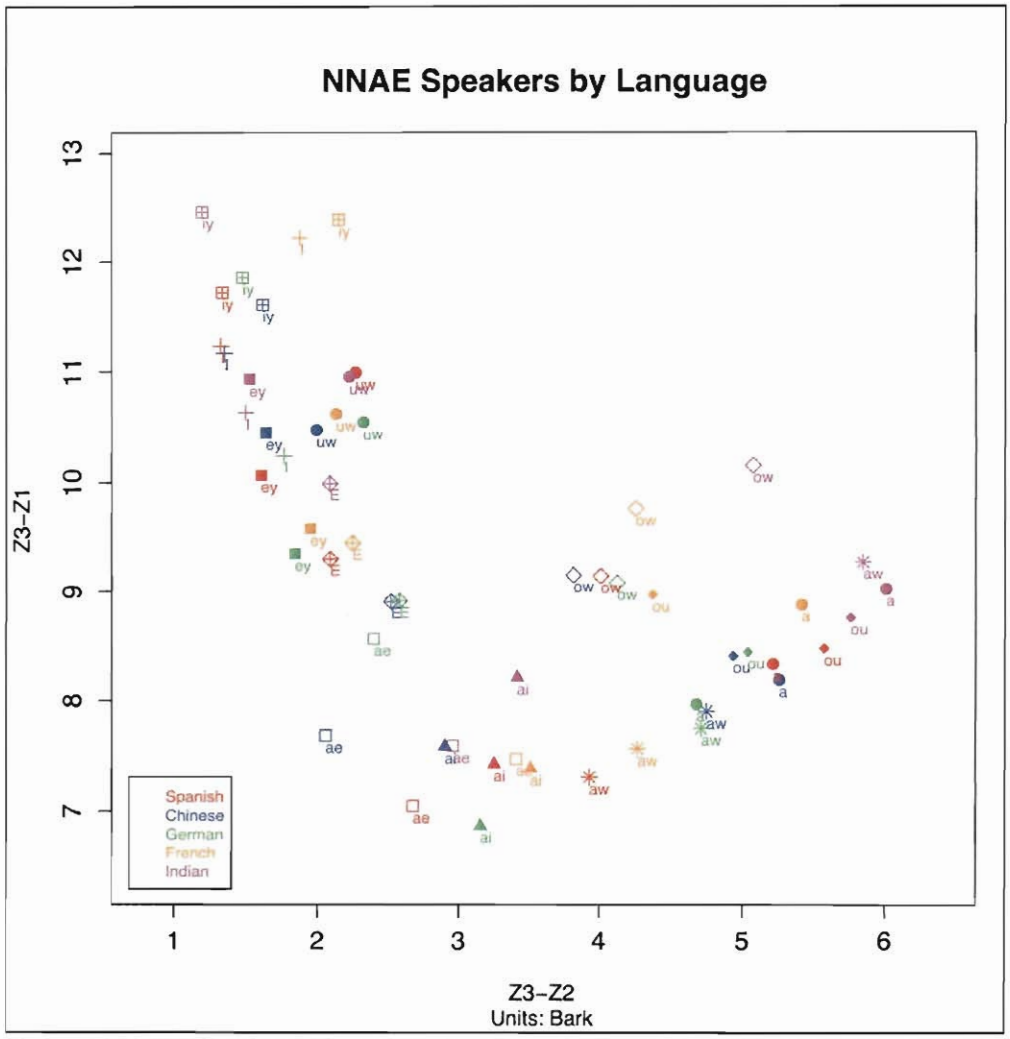


Figure 32 Normalized vowel space with means of NNAE speakers, by *language*. Note that the normalization procedure slightly skews the chart (for example, making [uw] appear fronter than in an unnormalized F1-F2 vowel space).

4.2.4 Conclusions

Several ANOVAs were used to test for significant differences among the NNAE speech. Testing for pitch revealed that neither mean pitch nor pitch range differed significantly by *language*.

The ANOVA for duration found that *vowel (type)* and the interaction of *vowel (type) x language* were significant. A clear difference between the length of

monophthongs and diphthongs exists particularly for German and French; the difference is much less pronounced for Spanish and Indian; and it is nonexistent for Chinese. The languages with less of a difference between monophthongs and diphthongs retained full glides on the diphthongs, so it was not the case that these speakers were monophthongizing the diphthongs. Nor was it the case that speech rate influenced the vowel length.

The ANOVA for formants revealed that the effects of *vowel* and *formant* were significant, as was the interaction of the two; these results were expected. *Language* was minimally important, reaching significance only in a three way interaction of *vowel x formant x language.*

The motivation for conducting these tests was to ascertain if the NNAE accents differed significantly along language lines. While differences do exist in the features analyzed here, they are not large. This indicates that other features may be more important in differentiating between these specific accents. However, this allows us to combine the NNAE results as a whole to compare to the AE results.

4.3 Comparing AE and NNAE voices

This section compares the data from the AE speakers in their three guises and the NNAE speakers. Like the previous two sections, it surveys pitch findings, vowel duration findings, and vowel formant findings.

4.3.1 Pitch

The average pitch for the regular voice was lower than the imitations and authentic accents (see Figure 33). The authentic pitch was highest, and the mean free imitation pitch was closest to it. Pitch range was not compared, as it had previously not been significant for either group of speakers.

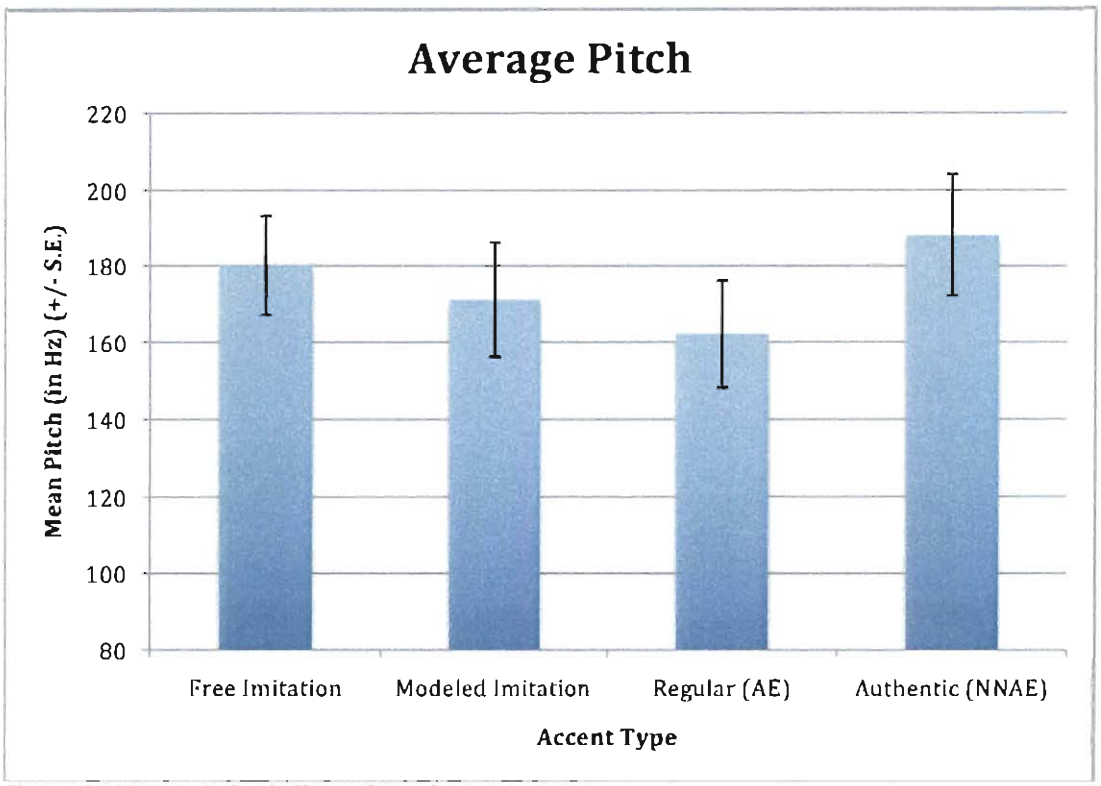


Figure 33 Mean pitch of all speakers, by *accent type*.

4.3.2 Duration

When comparing the mean duration of all four accent types, we can see in Figure 34 that overall the imitations are shorter than the regular voices, which are shorter than the authentic accents. This is one instance where the imitations are not more similar to authentic accents than the regular voices are.

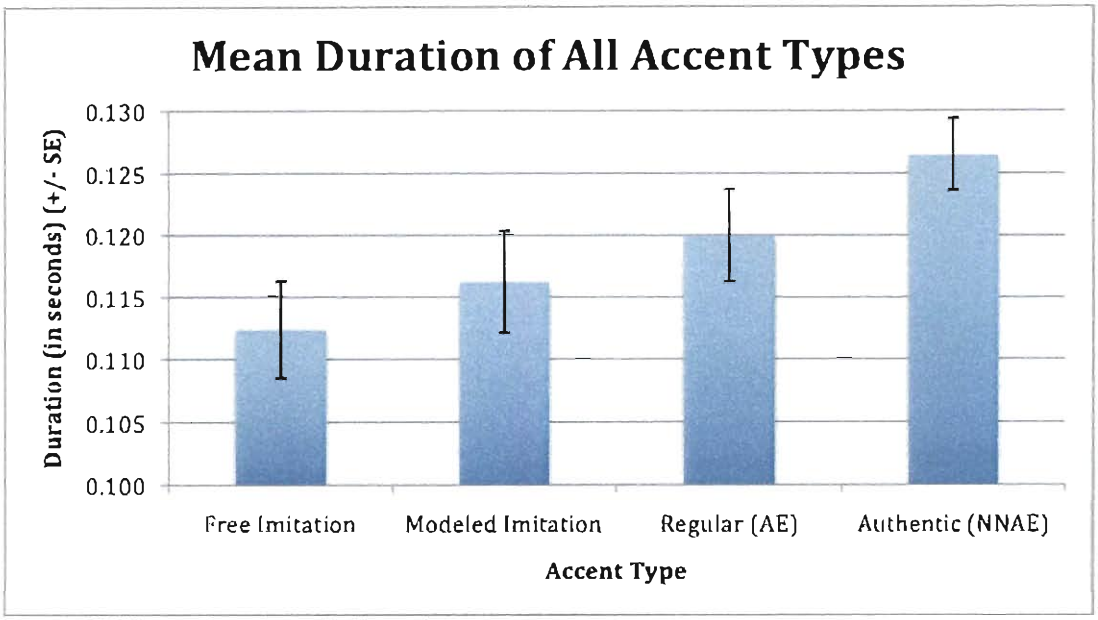


Figure 34 Means duration of all speakers, by *accent type*.

As seen in Figure 35, speech rate does not follow this pattern.

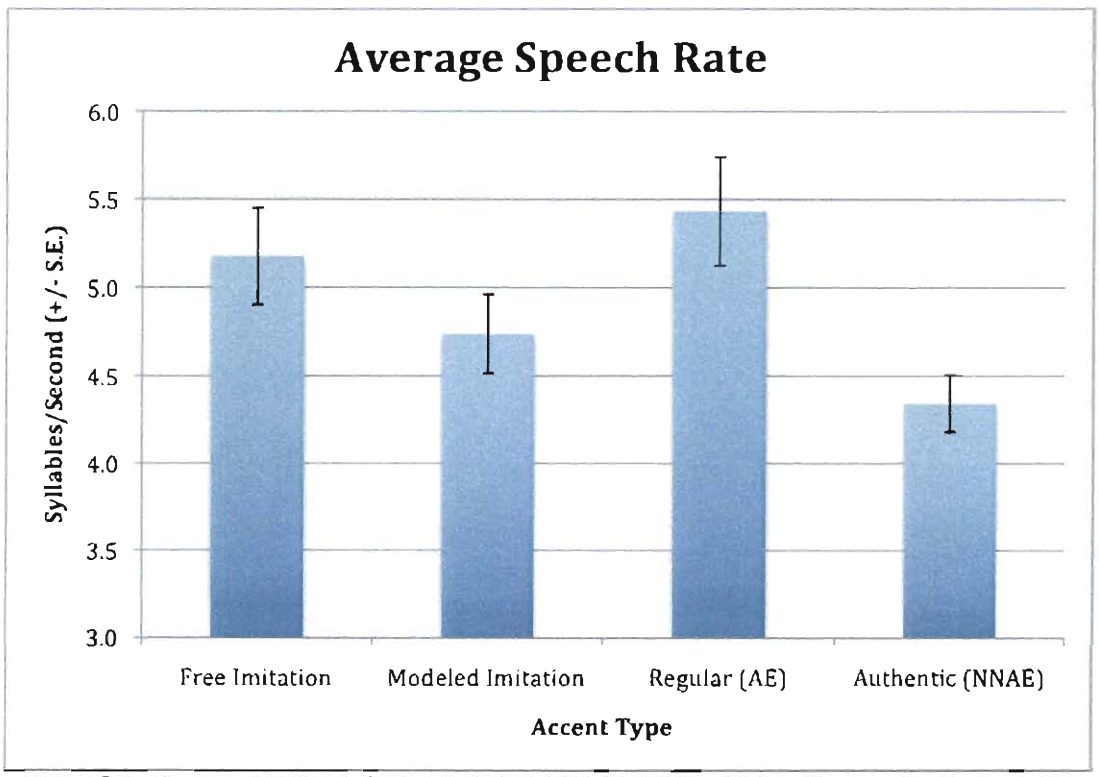


Figure 35 Mean speech rate of all speakers, by *accent type*. The higher the speech rate, the faster the speech.

This difference in duration is strange; why would the imitations, particularly the modeled imitation, have an overall shorter duration when the authentic accents, which served as the targets, have a longer duration? To begin to understand the durational variation, it is necessary to break down the duration by *vowel type* (see Figure 36). Here we can see that the NNAE speakers not only have longer vowel duration overall; they also have a larger difference between the monophthongs and the diphthongs than do the AE speakers in their regular guise. It seems that this large difference between vowel types may be reflected in the imitations; this is particularly the case with the modeled imitation, which has the largest difference between vowel types. Instead of creating this difference by lengthening the diphthongs, the imitations created this difference primarily by shortening the monophthongs.

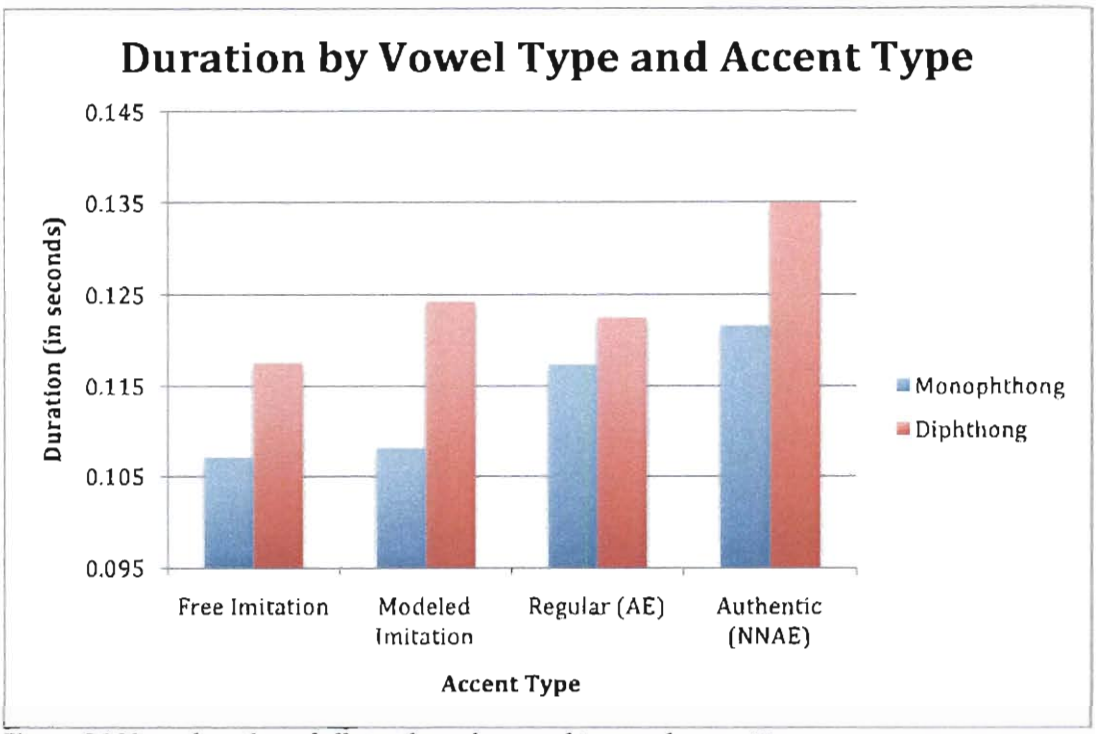


Figure 36 Mean duration of all speakers, by *vowel type* and *accent type*.

The issue with this conclusion is that for NNAE speakers the vowel length differs by *language*, and diphthongs are in fact not longer than monophthongs in all the languages. However, for AE imitations there is no significant difference by *language*. This may indicate that imitations base their vowel duration on broad artifacts as opposed to specific ones. For example, Figure 30 (reproduced as the "Authentic (NNAE)" column in Figure 36) is an average of the monophthong/diphthong difference for five languages; perhaps this is how the difference is represented within the broad artifact.

4.3.3 F1

The F1 of the regular voice was higher than that of the imitations and authentic speech (see Figure 37). These three accent types were remarkably similar, all averaging approximately 580 Hz.

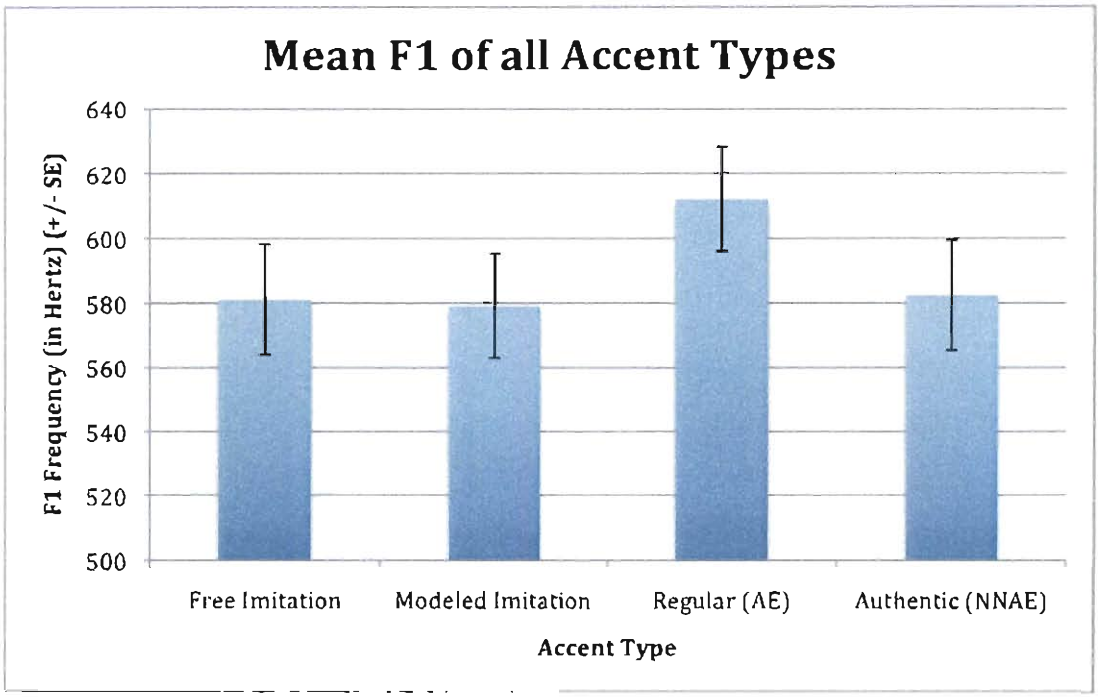


Figure 37 Mean F1 of all speakers, by *accent type*.

When mean F1 is broken down by both *accent type* and *language*, more trends are revealed (see Figure 38). F1 is approximately the same across all four accent types for Chinese. In French, German, and Indian, the regular accent has a much higher F1 than the imitations and the authentic accent. In these cases, imitations are more similar to the authentic accents than are the regular voices. For the Spanish accent, the authentic F1 is much higher than all of the AE voices.

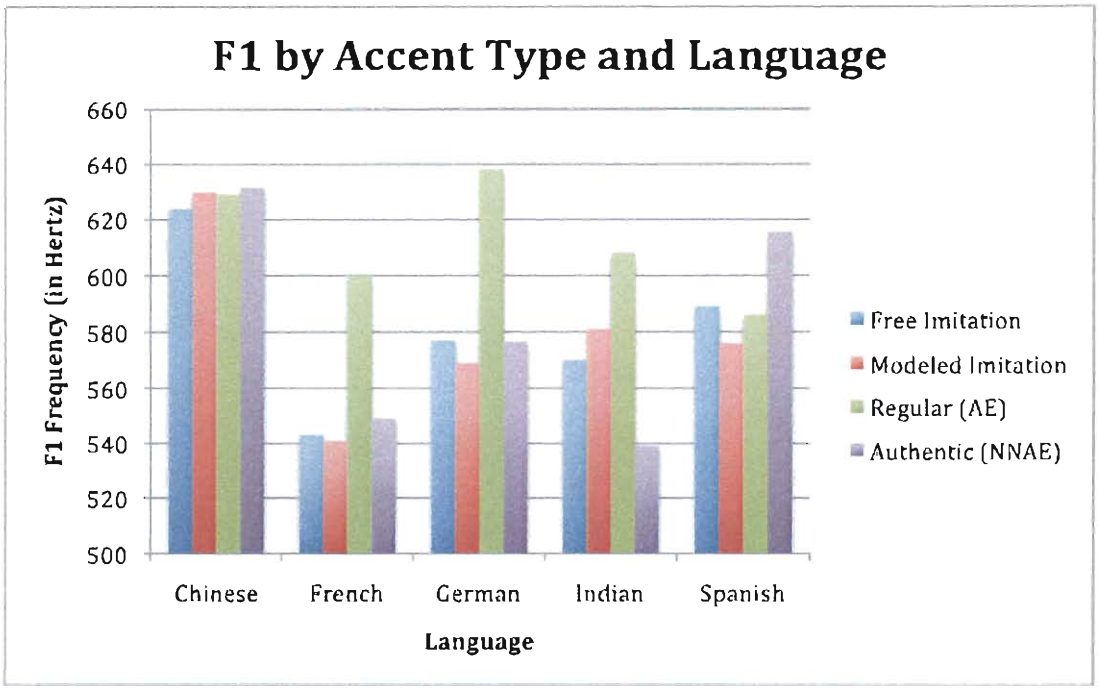


Figure 38 Mean F1 of all speakers, by *accent type* and *language*.

4.3.4 F2

On average, the F2 of the regular voice was higher than that of the imitations and authentic speech (see Figure 39). The mean modeled imitation was most similar to the mean authentic accent.

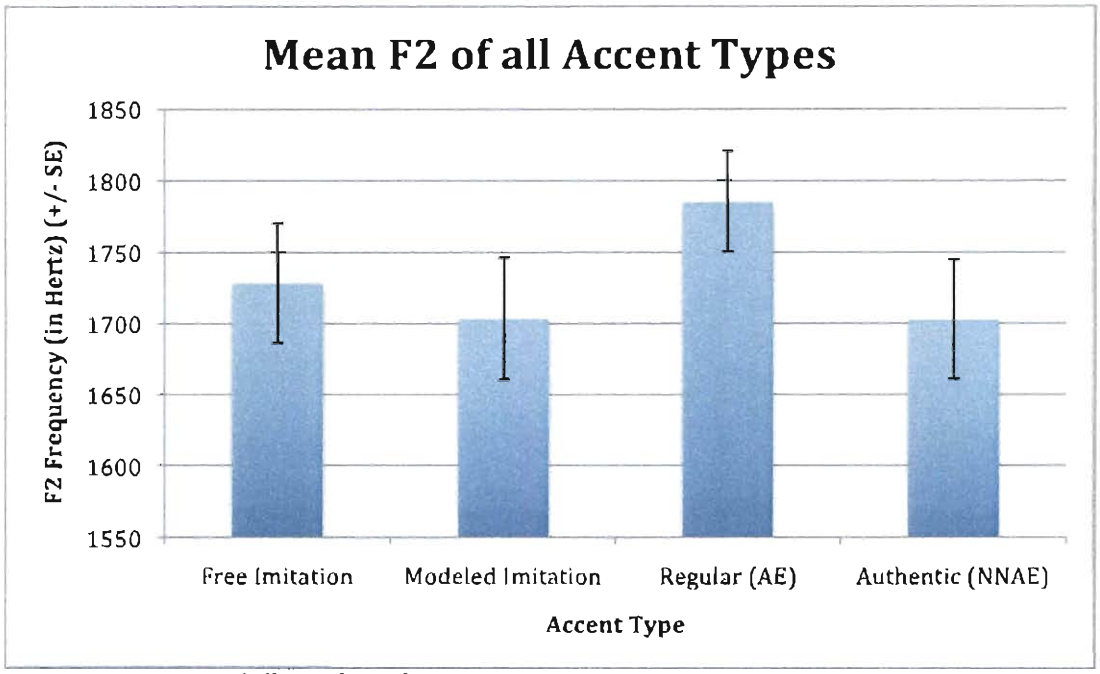


Figure 39 Mean F2 of all speakers, by *accent type*.

Figure 40 breaks down F2 means by *accent type* and *language*. For German, Indian, and Spanish, the mean F2 of the regular guise is noticeably higher than that of the imitations and authentic accent; in these cases the imitations are more similar to the authentic accents than to the regular guise. In the Chinese accent, the authentic and regular F2s are much higher than the imitations. In the French accent, all of the AE F2s are fairly high together, unlike the relatively low F2 of the authentic accent.

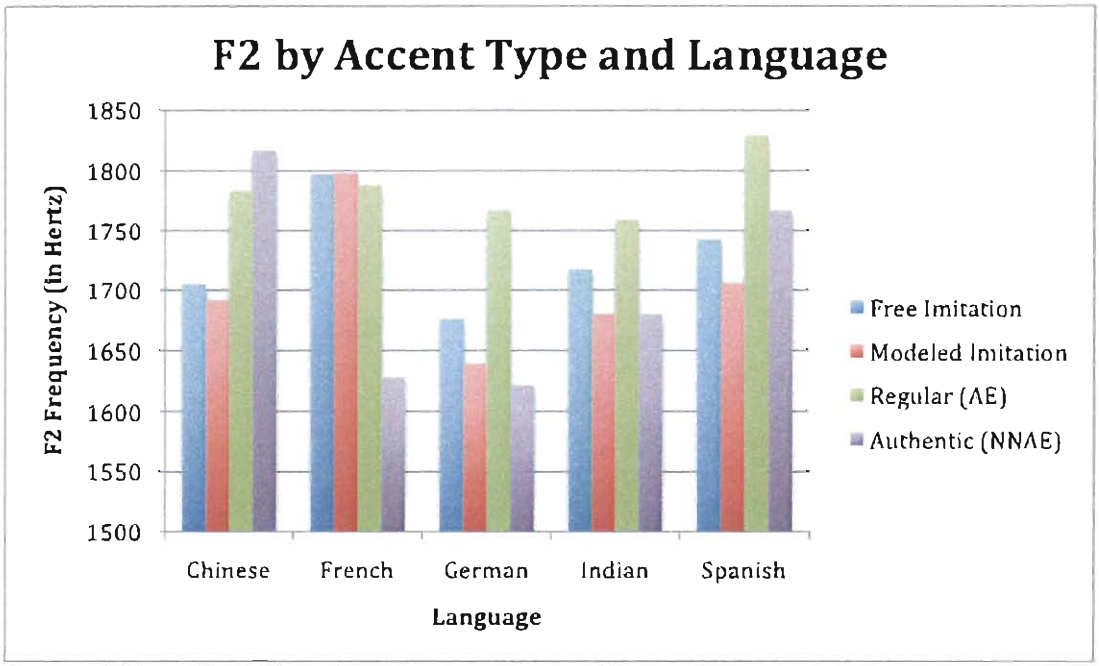


Figure 40 Mean F2 of all speakers, by *accent type* and *language*.

4.3.5 Vowel Space

A normalized vowel space of all speakers, separated by accent type, is shown in Figure 41.[17] In general, the imitations are closer to the authentic accents than to the regular guises. In several cases, the imitations are further from the regular guise than the authentic accent is, perhaps exaggerating a perceived difference that exists within a folk artifact. There is no consistent difference between the two imitation types whereby one of them is consistently nearer to the authentic target than the other.

[17] A normalized vowel space of all speakers separated by language is included in Appendix C.

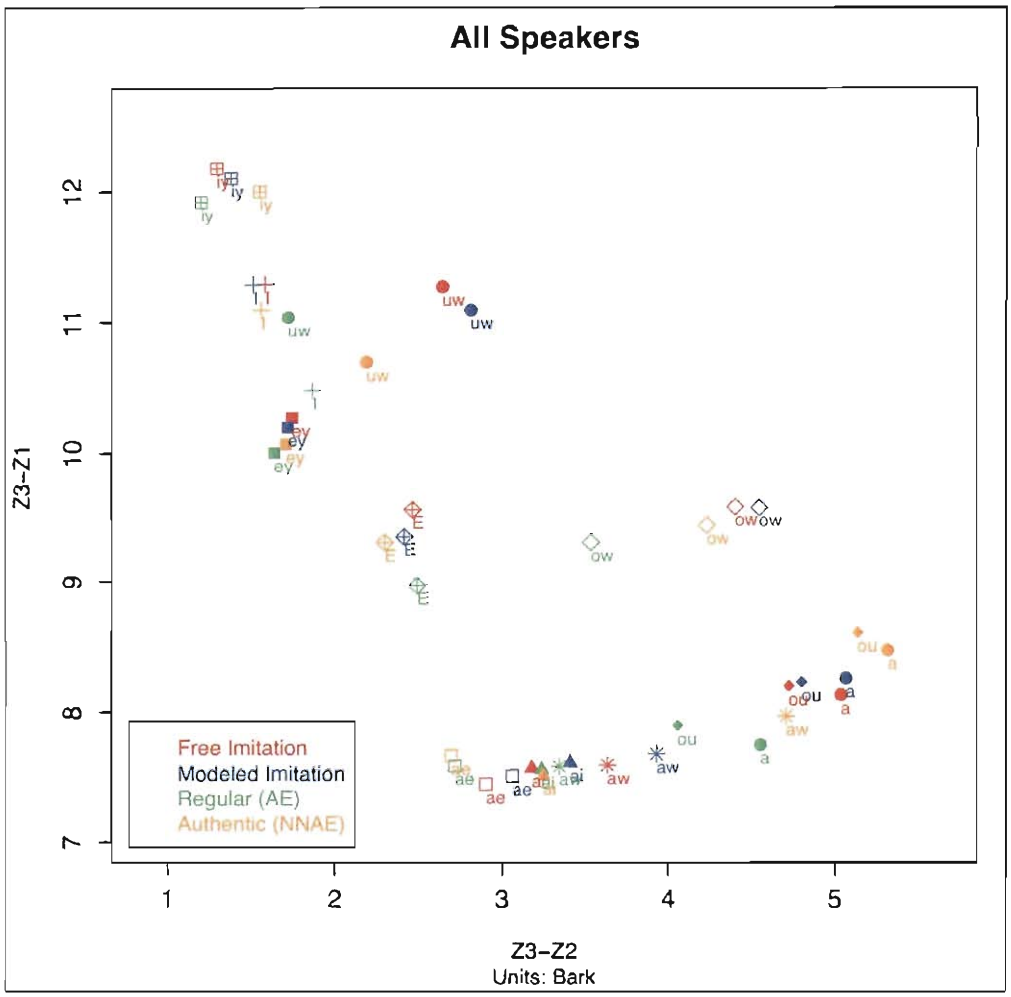


Figure 41 Normalized vowel space with nuclei means of all speakers, by *accent type*.

4.3.6 Conclusions

When we combine the AE and NNAE speech, we learn that imitations are more similar to authentic accents than to regular AE speech. While the modeled imitation is more similar to authentic accents in a comparison of average F2, the free imitation is more similar to authentic accents in a comparison of average pitch, and it is not possible to say for the features analyzed here whether the free or modeled imitations are more similar to the authentic accents.

4.4 Summary of Speaker Results

This chapter analyzed the phonetic features of four types of speech: free imitations, modeled imitations, regular (AE) voices, and authentic (NNAE) accents. This was done to address the first hypothesis: that speakers' modeled imitations would be more similar to the authentic accents than their free imitations. This hypothesis actually consists of two questions: Are free and modeled imitations significantly different from each other? If so, is one imitation more similar to authentic NNAE accents than the other?

In response to the first question, the linguistic features analyzed here (mean pitch, vowel duration, and vowel formants) did not show a consistent difference between free and modeled imitations. While the imitations were generally significantly different from speakers' regular voices, they were not significantly different from each other. It is therefore impossible to answer the second question and state whether one type of imitation is more similar to authentic accents, and the hypothesis is thus not supported.

The rationale behind the hypothesis was as follows: If free imitations are based on folk artifacts, and modeled imitations expand on these artifacts with features heard in authentic accents, then the phonetic details of the modeled imitations would be more similar to the authentic accents. Since the modeled imitations were not significantly different from the free imitations overall, we may assume that the vowel features in both imitations are strongly based on folk artifacts. It is important to remember that there are numerous other linguistic

features which were modified in addition to those assessed here, and it is plausible that these features are produced differently in the two imitation types. However, for the features of pitch, vowel duration, and vowel formants, there is no difference between imitation types and both are therefore predominantly influenced by folk artifacts, as opposed to authentic targets.

Chapter 5 Listener Results

This chapter presents the results from the listener's responses, centering on their identification of the speakers' language and authenticity. The three hypotheses regarding listener responses were as follows:

- Listeners will be able to identify the NNAE accents, both authentic and imitated.
- Listeners will not be able to distinguish between authentic and imitated accents.
- Listeners will react differently to free imitations and modeled imitations.

In section 5.1 I show that listeners can identify the accents they hear, verifying the first hypothesis above. In 5.2 I find that listeners are able to identify the authenticity of the accents they hear, refuting the second hypothesis above. The results related to the third hypothesis are mixed, as I show in 5.3, with free and modeled imitations identified differently for accent, but not for authenticity. Additional questions are addressed in 5.4, where I demonstrate that accent identification varies by language, that identification of accent and identification of authenticity are weakly correlated, that linguist judgments correlate with listener judgments, that listener confidence is correlated with accuracy, and that listener familiarity is minimally related to accuracy. I summarize the chapter in section 5.5.

5.1 Do listeners identify the accents they hear?

This question addresses the hypothesis that listeners will be able to identify the language of the accents. On average, listeners correctly identified the accent 70% of the time (pure chance was 20%).

5.1.1 Overall

For the initial analysis, the data was grouped in several different ways (see below). First, all the data was taken as a whole (A). Then the data was subdivided by group (B), where Group I heard the authentic accents and free imitations and Group II heard the authentic accents and modeled imitations. Next, responses were divided by the language (accent) of the speaker (C). Finally, the data was divided by both group and accent (D).

A. All listeners for all languages combined
B. Listeners by group for all languages combined
    1. Group I (Authentic and Free)
    2. Group II (Authentic and Modeled)
C. All listeners for each language
    1. French
    2. German
    3. Indian
    4. Chinese
    5. Spanish
D. Listeners by group for each language
    1. Group I
        i. French
        ii. German
        iii. Indian
        iv. Chinese
        v. Spanish
    2. Group II

i. French
ii. German
iii. Indian
iv. Chinese
v. Spanish

Multiple two-sided tests of proportions were then performed on each group. Table 23 below shows the results of these tests, the associated p-values, and the 95% confidence interval for the proportion of responses correct.

	Proportion correct	95% confidence interval		P-value for test
All listeners for all languages combined	0.706	0.677	0.733	0.00000
Group 1 listeners for all languages	0.744	0.705	0.780	0.00000
Group 2 listeners for all languages	0.664	0.620	0.705	0.00000
All listeners for French	0.510	0.440	0.579	0.83522
All listeners for German	0.793	0.731	0.845	0.00000
All listeners for Indian	0.673	0.604	0.735	0.00000
All listeners for Chinese	0.798	0.736	0.849	0.00000
All listeners for Spanish	0.755	0.690	0.810	0.00000
Group 1 listeners for French	0.546	0.448	0.641	0.38648
Group 1 listeners for German	0.796	0.706	0.865	0.00000
Group 1 listeners for Indian	0.704	0.607	0.786	0.00004
Group 1 listeners for Chinese	0.880	0.799	0.932	0.00000
Group 1 listeners for Spanish	0.796	0.706	0.865	0.00000
Group 2 listeners for French	0.470	0.370	0.572	0.61708
Group 2 listeners for German	0.790	0.695	0.862	0.00000
Group 2 listeners for Indian	0.640	0.537	0.732	0.00693
Group 2 listeners for Chinese	0.710	0.609	0.794	0.00004
Group 2 listeners for Spanish	0.710	0.609	0.794	0.00004

Table 23 Results from multiple two-sided tests of proportions for accent identification.

We see that, on the whole, listeners can indeed identify the accents at a significant rate (p = .000). This holds across groups and across most languages. The only instances where listeners cannot identify the accents at a significant rate are for the French speakers. This is likely due to the specific NNAE female speaker, who listeners only identified as French 35% of the time (see Table 24).

Speaker	Accent Type	Accent Identification
F16 (male)	Authentic	73%
F21 (female)	Authentic	35%
A28 (male)	Imitation – Free	87%
	Imitation – Modeled	80%
A09 (female)	Imitation – Free	88%
	Imitation – Modeled	48%
All	French Accents	65%

Table 24 Accent identification of individual French speakers.

5.1.2 Imitated and Authentic Voices

The accents of imitated voices were identified 69% of the time; the accents of authentic voices were identified 72% of the time. There is very little difference between these rates. However, since the previous testing did not differentiate between imitated and authentic voices, an additional logistic regression was used to determine whether the accent identification of the two was significantly different. For each logit regression, the model aimed to predict the correctness of accent identification from the type of voice heard (authentic or imitation), and from the participant's group (I or II). The tables below show the parameter estimates, as well as the results of significance tests on the model. Table 25 reveals that the model is in fact significant. Table 26 shows the significance of each individual predictor. Table 27 presents the odds ratio estimates.

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	9.0079	2	0.0111
Score	8.998	2	0.0111
Wald	8.9448	2	0.0114

Table 25 Significance of the model for accent identification of imitated and authentic voices.

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard error**	**Wald Chi-square**	**Pr > ChiSq**
Real	1	-0.1307	0.1367	0.9141	0.339
Group	1	-0.3884	0.1368	8.0589	0.0045

Table 26 Analysis of Maximum Likelihood Estimates for accent identification of imitated and authentic voices.

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
Real	0.877	0.671	1.147
Group	0.678	0.519	0.887

Table 27 Odds Ratio Estimates for accent identification of imitated and authentic voices.

The type of voice heard (the *real* parameter) was not significant in the regression, which implies that there is no significant difference between listeners' identification of imitation and of authentic voices. Listeners are just as accurate at identifying the accent of imitations as they are at identifying the accent of authentic voices.

The *group* parameter was significant ($p = .005$), and the odds ratio tells us that listeners in Group II (those who heard authentic accents and modeled imitations) were about 0.7 times more likely to identify the accents of imitations than were listeners from Group I (those who heard authentic accents and free imitations); i.e., Group II listeners were 30% less likely to identify the speakers' accents. The significance of this group variable indicates that there is likely a significant difference between the way listeners identified the free and the modeled imitations. This is further addressed in 5.3.1.

5.2 Do listeners identify the authenticity of the accents?

This question springs from the second hypothesis, which was that listeners would not be able to tell whether an accent was authentic or imitated. On the whole, listeners were able to correctly identify the authenticity 65% of the time (versus pure chance of 50%).

5.2.1 Overall

This question is essentially the same as the previous one, except the data concerned is authenticity, instead of accent. The same groupings were used and multiple two-sided tests of proportions were again performed on each group. Table 28 presents the results of these tests, the associated p-values, and the 95% confidence interval for the proportion of responses correct.

	Proportion correct	95% confidence interval		P-value for test
All listeners for all languages combined	0.651	0.621	0.680	0.00000
Group 1 listeners for all languages	0.674	0.632	0.713	0.00000
Group 2 listeners for all languages	0.626	0.582	0.668	0.00000
All listeners for French	0.654	0.584	0.717	0.00001
All listeners for German	0.567	0.497	0.635	0.06119
All listeners for Indian	0.663	0.594	0.726	0.00000
All listeners for Chinese	0.716	0.649	0.775	0.00000
All listeners for Spanish	0.654	0.584	0.717	0.00001
Group 1 listeners for French	0.667	0.569	0.753	0.00076
Group 1 listeners for German	0.519	0.421	0.615	0.77283
Group 1 listeners for Indian	0.657	0.559	0.744	0.00150
Group 1 listeners for Chinese	0.806	0.716	0.873	0.00000
Group 1 listeners for Spanish	0.722	0.626	0.802	0.00001
Group 2 listeners for French	0.640	0.537	0.732	0.00693
Group 2 listeners for German	0.620	0.517	0.714	0.02145
Group 2 listeners for Indian	0.670	0.568	0.759	0.00097
Group 2 listeners for Chinese	0.620	0.517	0.714	0.02145
Group 2 listeners for Spanish	0.580	0.477	0.677	0.13361

Table 28 Results from multiple two-sided tests of proportions for authenticity identification.

Listeners are able to identify the authenticity of the voices they heard at a significant rate (p = .000). This significance held across groups and across most languages (though listeners in Group I were unable to identify the authenticity of the German speakers, and listeners in Group II were unable to identify the authenticity of the Spanish speakers). Overall, listeners can recognize whether an accent is real or fake. However, while there is positive correlation in authenticity, this correlation is not highly impressive.

5.2.2 Imitated and Authentic Voices

The authenticity of imitated voices was identified 65% of the time; the authenticity of authentic voices was identified 67% of the time. There is again very little difference between these rates. However, as in 5.1.2, a logistic regression was run to ensure that the authenticity identification of imitated and authentic voices was not statistically different.

The model itself was not significant. The type of voice heard (*real*) was not significant; therefore, listeners did not identify the authenticity of imitated and authentic voices differently. In other words, listeners are just as accurate at identifying the authenticity of imitations as they are at identifying the authenticity of authentic voices. The *group* variable was also not significant, suggesting there was not a difference in the authenticity identification of free and modeled speakers. This specific issue is addressed in the next section.

5.3 Do listeners identify the free imitations and modeled imitations differently?

The third hypothesis is that listeners will rate the free and modeled imitations differently. While the *group* results in the previous sections hinted at the answer to this, the imitations were lumped together with the authentic voices, and it was necessary to do further tests on the imitations alone to determine whether there was in fact a significant difference between the identification of free and

modeled imitations. Figure 42 presents the rates of correct identification of accent and authenticity for both types of imitations.

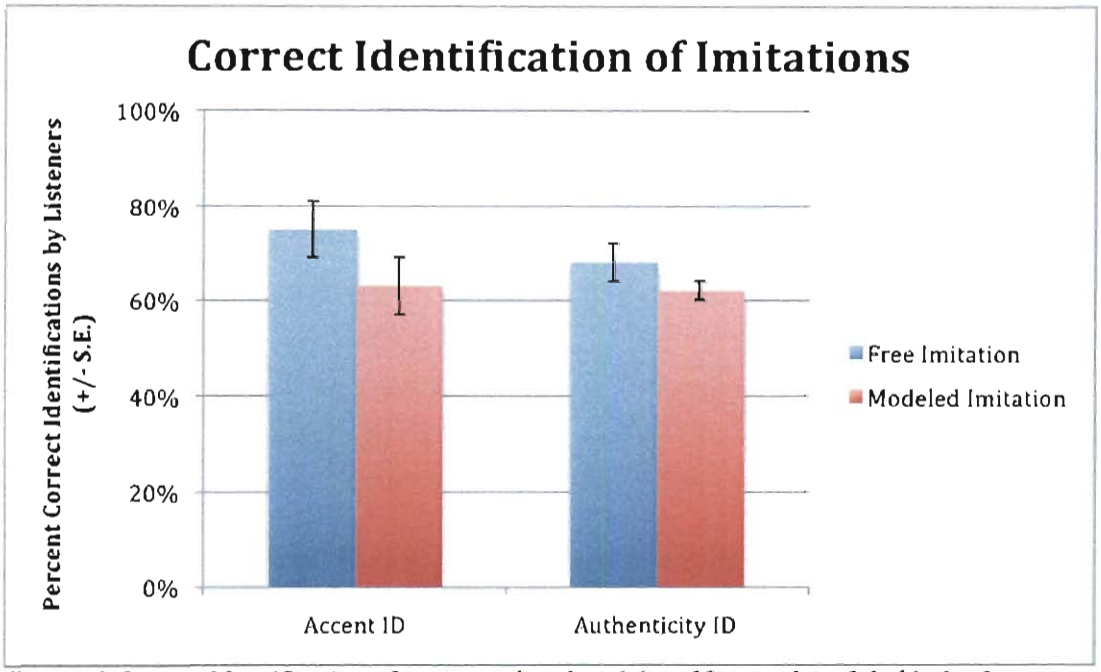


Figure 42 Correct identification of accent and authenticity of free and modeled imitations.

Two logistic regressions were implemented to investigate whether there was a significant difference between the identification of free and modeled imitations, one for accent identification and one for authenticity identification. Each logit regression model aimed to predict the correctness of identification from the type of imitation heard (*group*[18]).

[18] Note that in this case, *group* refers only to listeners' responses to imitations, and not to the responses for both authentic and imitated accents, as was the case in the previous tests.

5.3.1 Accent Identification of Free and Modeled Imitations

Listeners correctly identified the accent of free imitations 75% of the time; they correctly identified the accent of modeled imitations 63% of the time. There is a significant difference between these rates of accent identification ($p = .002$). In particular, listeners are 45% less likely to identify the accents of modeled imitations than they are to identify the accents of free imitations. Results from the logistic regression are shown in Table 29, Table 30, and Table 31 below.

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	9.3683	1	0.0022
Score	9.3474	1	0.0022
Wald	9.2502	1	0.0024

Table 29 Significance of the model for accent identification of free and modeled imitations.

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Sqaure**	**Pr > ChiSq**
Group	1	-0.5848	0.1923	9.2502	0.0024

Table 30 Analysis of Maximum Likelihood Estimates for accent identification of free and modeled imitations.

Odds Ratio Estimates			
Comparison	**Odds ratio**	**95% Wald Confidence Limits**	
Group	0.557	0.382	0.812

Table 31 Odds Ratio Estimates for accent identification of free and modeled imitations.

5.3.2 Authenticity Identification of Free and Modeled Imitations

Listeners correctly identified the authenticity of free imitations 68% of the time; they correctly identified the authenticity of modeled imitations 62% of the time. The logistic regression model for authenticity identification was not

significant, which implies that there was no significant difference between these rates of authenticity identification of free and modeled imitations.

Overall, we see that listeners are better at identifying the accent of free imitations than modeled imitations. However, there is not a significant difference in the authenticity identification of the two imitation types.

5.4 Additional Questions

I was able to use this data to address several other questions. These included whether identification varied by language (5.4.1), whether the identification of accent and the identification of authenticity were related (5.4.2), whether the linguists' judgments were correlated with the listeners' judgments (5.4.3), whether listener accuracy was related to listener confidence (5.4.4), and whether listener accuracy was related to listener familiarity (5.4.5). Where correlations are discussed, both Pearson and Spearman tests were conducted; results were consistent between the two each time.

5.4.1 Identification by Language

The first additional questions were if identification varied by language and whether there were common language misidentifications. Figure 43 shows the correct identifications by language. According to this chart, the accents most likely to be identified were Chinese and German, followed by Spanish, then Indian, and

behind that French. The authenticity was most likely to be correctly identified for Chinese speakers, followed by Indian and Spanish and French, and then German.

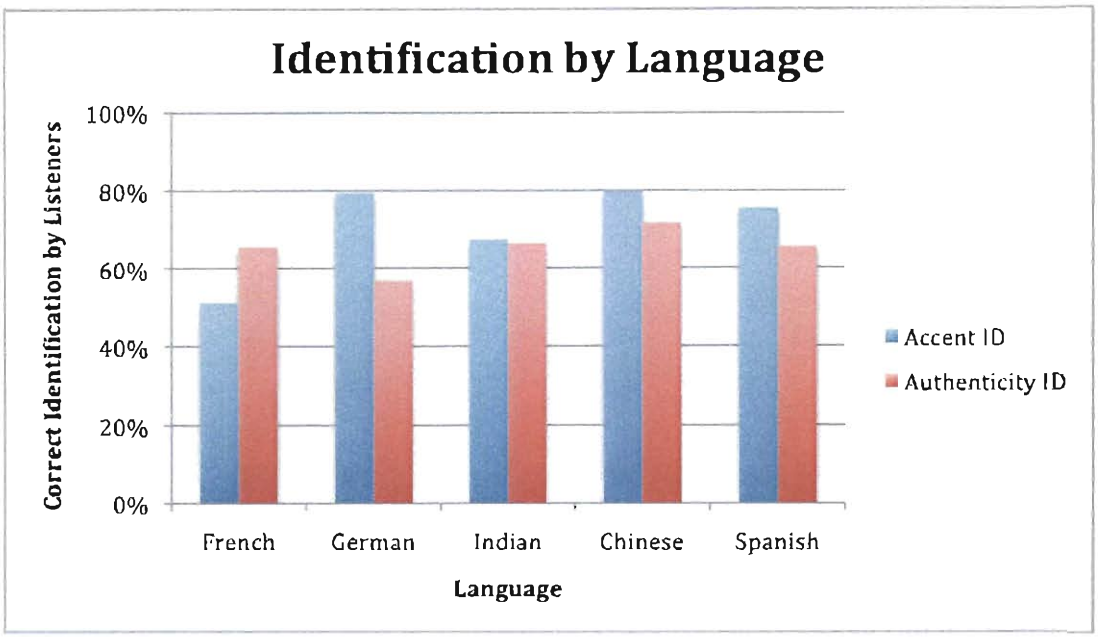


Figure 43 Identification of all speakers, by language.

A logistic regression was performed with correctness as the dependent variable and language as the independent variable. It found that accent identification did vary significantly by language (see Table 32 and Table 33), though authenticity identification did not. In order to compare the accent identification of languages, five sets of odds ratios were obtained, each using a different language as a reference; Table 34 shows these odds ratios and their respective confidence intervals for all possible pairs of languages. (The odds ratio for the comparisons *A vs B* can be interpreted as follows: A listener is *x* times more likely to identify an accent from language *A* than from language *B*, where *x* is the odds ratio. So for

example, the first line of Table 34 would be read: "A listener is 3.692 times more likely to identify a German accent than a French accent.")

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	56.1188	4	<.0001
Score	58.2223	4	<.0001
Wald	55.4004	4	<.0001

Table 32 Significance of the model for identification by language.

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard error**	**Wald Chi-square**	**Pr > ChiSq**
GERMAN	1	1.3063	0.2204	35.1434	<.0001
INDIAN	1	0.6837	0.2027	11.376	0.0007
CHINESE	1	1.3359	0.2215	36.3653	<.0001
SPANISH	1	1.086	0.2126	26.0819	<.0001

Table 33 Analysis of Maximum Likelihood Estimates for identification by language.

Odds Ratio Estimates				
Comparison	**Odds ratio**	**95% Wald Confidence Interval**		**REFERENCE**
GERMAN vs FRENCH	3.692	2.397	5.687	FRENCH
INDIAN vs FRENCH	1.981	1.332	2.947	
CHINESE vs FRENCH	3.803	2.464	5.871	
SPANISH vs FRENCH	2.962	1.953	4.494	
FRENCH vs GERMAN	0.271	0.176	0.417	GERMAN
INDIAN vs GERMAN	0.537	0.344	0.836	
CHINESE vs GERMAN	1.03	0.639	1.659	
SPANISH vs GERMAN	0.802	0.506	1.272	
FRENCH vs INDIAN	0.505	0.339	0.751	INDIAN
GERMAN vs INDIAN	1.864	1.196	2.904	
CHINESE vs INDIAN	1.92	1.23	2.997	
SPANISH vs INDIAN	1.495	0.974	2.295	
FRENCH vs CHINESE	0.263	0.17	0.406	CHINESE
GERMAN vs CHINESE	0.971	0.603	1.564	
INDIAN vs CHINESE	0.521	0.334	0.813	
SPANISH vs CHINESE	0.779	0.49	1.238	
FRENCH vs SPANISH	0.338	0.223	0.512	SPANISH
GERMAN vs SPANISH	1.246	0.786	1.976	
INDIAN vs SPANISH	0.669	0.436	1.027	
CHINESE vs SPANISH	1.284	0.808	2.04	

Table 34 Odds Ratio Estimates for identification by language.

There was a significant difference by language for accent identification, but not for authenticity identification. The accents in order from most to least often identified are Chinese, German, Spanish, Indian, French.

In order to determine whether some languages were commonly mistaken for others, Figure 44 plots the listener responses below. We can see that French accents were often mistaken as Spanish accents; German accents were very often mistaken as Chinese accents; Indian accents were often mistaken as German accents; Chinese

accents were very often mistaken as French accents; and Spanish accents were very often mistaken as Indian accents. In fact, the German and Spanish accents were more often misidentified then correctly identified. These results are surprising, particularly because none of the misidentifications were complementary (with each accent being perceived as the other).

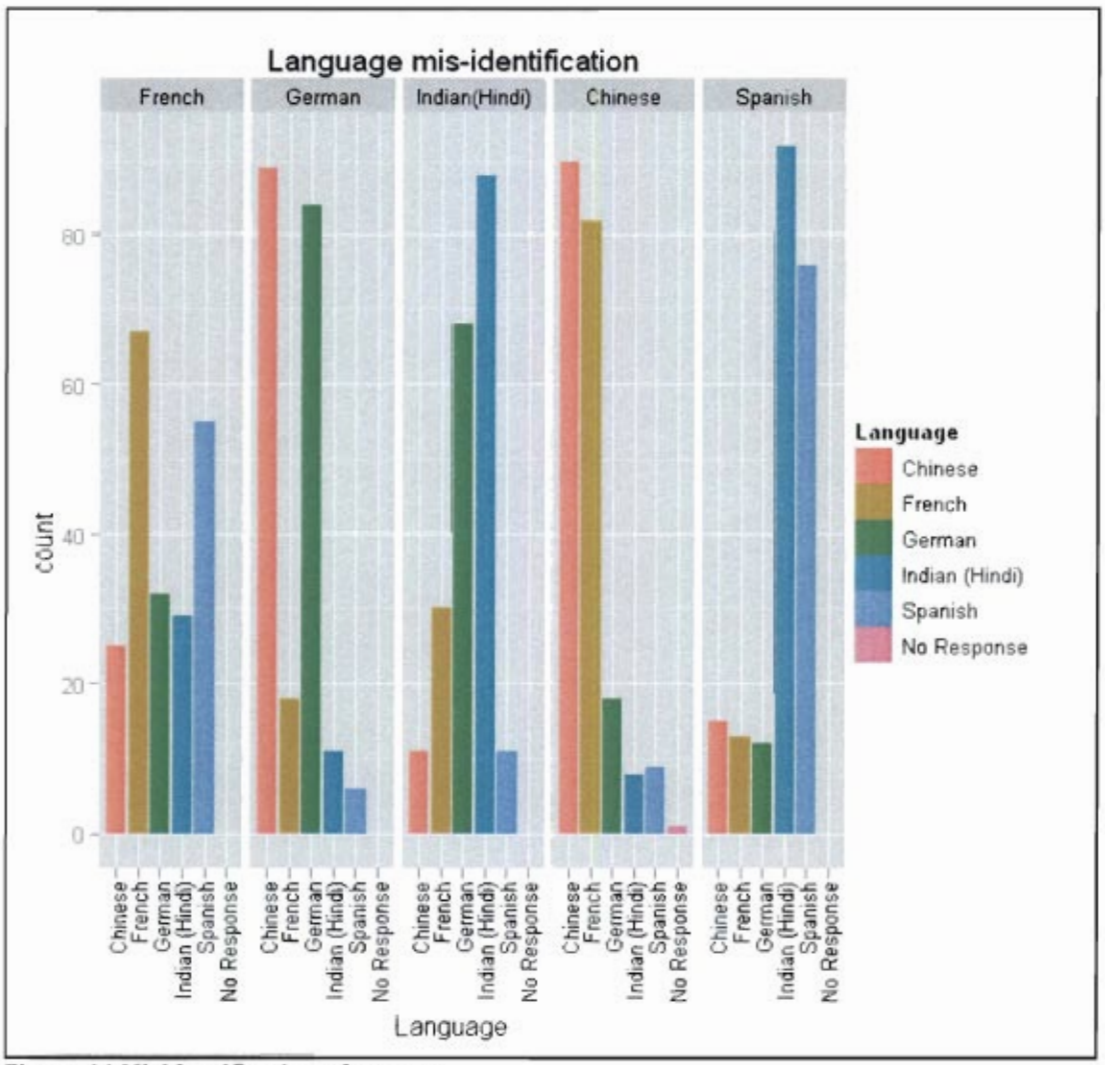


Figure 44 Misidentification of accents.

5.4.2 Identification of Accent and Authenticity

In order to determine whether the identifications of accents and the identifications of authenticity were related, a Pearson correlation was run. The identifications were correlated (r = + .072, n = 1040, p = .019, two-tailed). Spearman's test drew the same conclusion: there is a correlation, but the association is very weak, as depicted in Figure 45. If a listener gets one type of identification correct, they are slightly more likely to get the other identification correct as well.

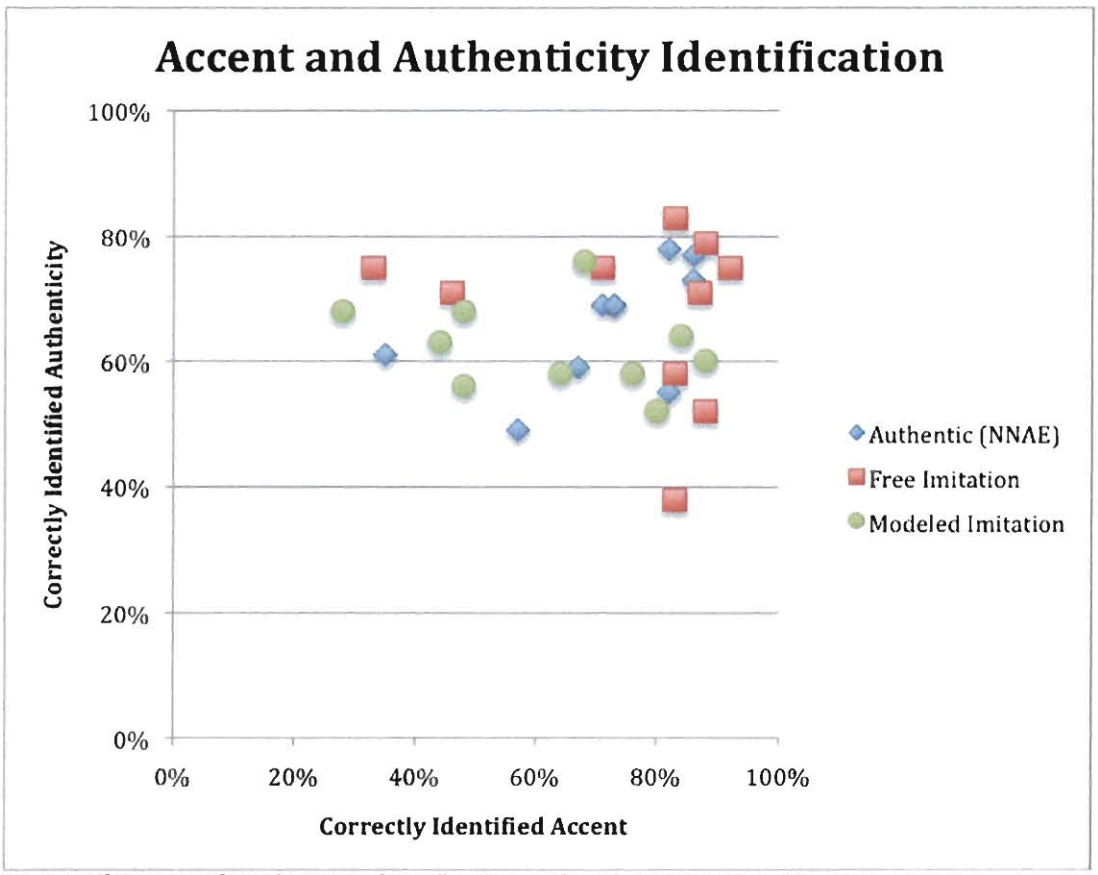


Figure 45 Scatter plot of accent identification and authenticity identification.

5.4.3 Linguist Judgments and Listener Identification

Another follow-up question was whether the original linguist judgments correlated with listener judgments. As noted in 3.3.1 and 3.3.2, linguists judged each NNAE speaker for how representative their accent was, and each AE imitation for how believable their accent was. A Pearson correlation revealed a significant relationship for accent identification between the listener responses and linguist ratings ($r = +.504$, $n = 30$, $p = .005$, two-tailed). The higher the linguists rated a speaker, the more likely that listeners correctly identified a speaker's accent (see Figure 46).

There was not a significant correlation for authenticity between listeners' and linguists' judgments. This is interesting in light of the fact that the linguists' judgments of imitations were regarding their believability, a measurement which was assumed to be closely related to authenticity.

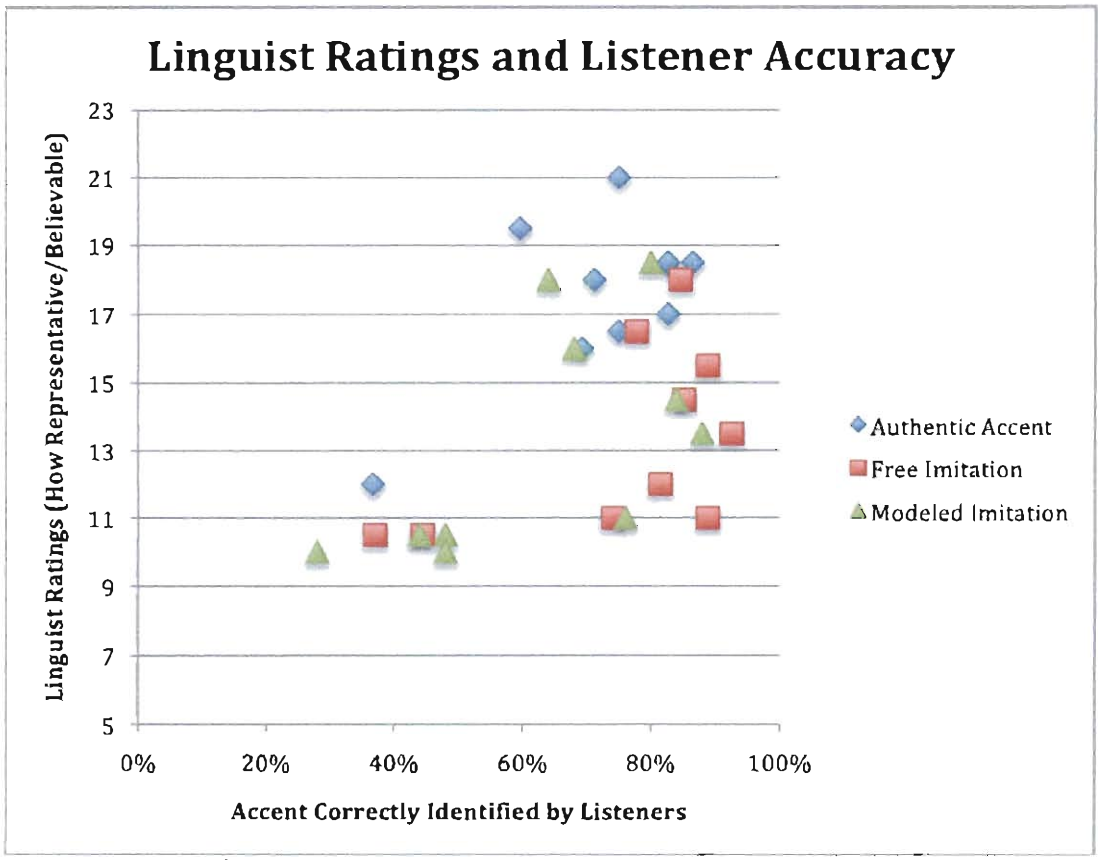


Figure 46 Scatter plot of linguist ratings and listener accuracy at identifying the accent. Note that authentic accents (NNAE) were rated on how representative they were, while imitations (AE) were rated on how believable they were.

5.4.4 Listener Accuracy and Confidence

Listeners rated their confidence in each identification they made, where 1 was "Very certain" and 5 was "Very uncertain." Both Pearson and Spearman revealed significant correlations for accent identification ($r = -.35$, $n = 1040$, $p = .000$, two-tailed) and for authenticity identification ($r = -.11$, $n = 1040$, $p = .017$, two-tailed). This means that there is a slightly negative correlation between identification and confidence: the less certain a listener is in their judgment, the less likely they are to be correct. (The negative is a result of the manner of rating. We

may also say that the more confident a listener is in their identification, the more likely they are to be correct.)

Since there was a relationship between identification and confidence, two logistic regressions were performed to determine whether confidence could predict accuracy of accent and authenticity identification accuracy. Results from the logit regression for accent identification are shown in Table 35 and Table 36. It was found that confidence can predict accuracy of accent identification: the greater the listener's certainty in their decision, the more likely they were to accurately identify the accent. In particular, as Table 37 shows, a listener with a confidence level of 1 ("Very certain") is 13 times more likely to correctly identify the accent than a listener with a confidence level of 5 ("Very uncertain").

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	145.2895	4	<.0001
Score	145.7904	4	<.0001
Wald	127.6309	4	<.0001

Table 35 Significance of the model for accent identification and confidence.

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard error**	**Wald chi-square**	**Pr > ChiSq**
CERTAINTY 1	1	1.4741	0.1506	95.8698	<.0001
CERTAINTY 2	1	0.5614	0.1251	20.1388	<.0001
CERTAINTY 3	1	-0.6926	0.1594	18.8776	<.0001
CERTAINTY 4	1	-0.2293	0.1634	1.9699	0.1605

Table 36 Analysis of Maximum Likelihood Estimates for accent identification and confidence.

Odds Ratio Estimates			
Comparison	**Odds ratio**	**95% Wald Confidence Interval**	
CERTAINTY 1 vs 5	13.299	7.366	24.01
CERTAINTY 2 vs 5	5.339	3.076	9.267
CERTAINTY 3 vs 5	1.523	0.832	2.791
CERTAINTY 4 vs 5	2.421	1.313	4.466

Table 37 Odds Ratio Estimates for accent identification and confidence.

Results from the logistic regression for authenticity identification are shown in Table 38 and Table 39. Once again, listener confidence can predict identification. The odds ratios for this model are lower than those for accent identification (see Table 40). In particular, a listener with a confidence level of 1 is about 2.4 times more likely to correctly recognize an accent's authenticity than one with a confidence level of 5.

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	12.6986	4	0.0128
Score	12.8845	4	0.0119
Wald	12.6901	4	0.0129

Table 38 Significance of the model for authenticity identification and confidence.

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard error**	**Wald chi-square**	**Pr > ChiSq**
CERTAINTY 1	1	0.4266	0.1625	6.8967	0.0086
CERTAINTY 2	1	0.2361	0.1115	4.4855	0.0342
CERTAINTY 3	1	-0.0304	0.1372	0.0489	0.8249
CERTAINTY 4	1	-0.1653	0.1654	0.9985	0.3177

Table 39 Analysis of Maximum Likelihood Estimates for authenticity identification and confidence.

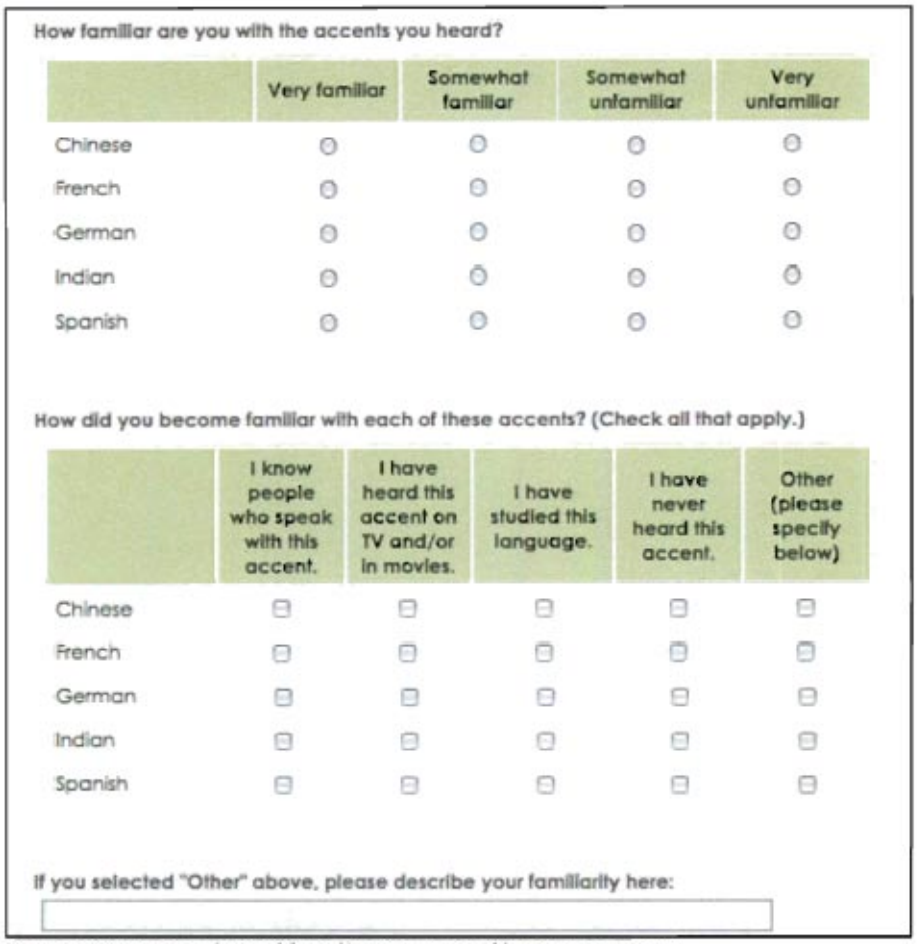
Odds Ratio Estimates			
Comparison	**Odds ratio**	**95% Wald Confidence Interval**	
CERTAINTY 1 vs 5	2.444	1.307	4.57
CERTAINTY 2 vs 5	2.02	1.166	3.501
CERTAINTY 3 vs 5	1.548	0.862	2.78
CERTAINTY 4 vs 5	1.352	0.72	2.541

Table 40 Odds Ratio Estimates for authenticity identification and confidence.

Overall, listeners' confidence did indeed predict their identification accuracy. This is interesting because the question of whether or not confidence can predict accuracy has produced very mixed results (e.g. Hammersley & Read 1996).

5.4.5 Listener Accuracy and Familiarity

The final test was to investigate whether listener accuracy was related to their familiarity with these accents. Listener familiarity was assessed in two ways (see Figure 47). First, listeners were asked to rate how familiar they were with each accent (*level*), on a scale of 1-4, with 1 being "Very familiar" and 4 being "Very unfamiliar." Then listeners were asked the manner in which they became familiar with each accent (*method*). The method options were Personal ("I know people who speak with this accent"), Media ("I have heard this accent on TV and/or in movies"), Study ("I have studied this language"), Never ("I have never heard this accent"), and Other; listeners could select all the methods that applied. Very few listeners selected Other, and the responses that did were reclassified as Personal based on their additional descriptions.



How familiar are you with the accents you heard?

	Very familiar	Somewhat familiar	Somewhat unfamiliar	Very unfamiliar
Chinese	○	○	○	○
French	○	○	○	○
German	○	○	○	○
Indian	○	○	○	○
Spanish	○	○	○	○

How did you become familiar with each of these accents? (Check all that apply.)

	I know people who speak with this accent.	I have heard this accent on TV and/or in movies.	I have studied this language.	I have never heard this accent.	Other (please specify below)
Chinese	☐	☐	☐	☐	☐
French	☐	☐	☐	☐	☐
German	☐	☐	☐	☐	☐
Indian	☐	☐	☐	☐	☐
Spanish	☐	☐	☐	☐	☐

If you selected "Other" above, please describe your familiarity here:

Figure 47 Screenshot of familiarity page of listener survey.

A two-tailed Pearson correlation showed that there was not a significant correlation between the level and method of familiarity. The two types of familiarity, level and method, were therefore not related.

A Pearson correlation also found that there was not a significant relationship between level of familiarity and accent identification. However, there was a significant correlation between level of familiarity and authenticity identification ($r = -.136$, $n = 255$, $p = .029$, two-tailed). This means that the less familiar a listener is with an accent, the less likely they are to correctly identify their authenticity. In

other words, the more familiar they are, the more likely they are to correctly identify whether a voice is authentic or imitated.

There were eight possible combinations of responses to *method* of familiarity.

1. Personal only
2. Personal and Media
3. Personal and Media and Study
4. Personal and Study
5. Media only
6. Media and Study
7. Study only
8. Never

Most listeners selected options 1, 2, or 5; their primary experience was through personal interaction and media exposure. The methods were tested for correlations with correct identifications. None of the methods reached significance, neither for accent identification nor for authenticity identification.

Overall, there was very little effect of familiarity on listener identification. The level of familiarity did influence the authenticity identification, but it was unrelated to accent identification, and method of familiarity was not related to any identification.

5.5 Summary of Listener Results

Listeners recognized the accents at a significant rate, which confirms the hypothesis that listeners would be able to identify NNAE accents. The strong folk artifact status of these accents may have assisted in this recognition. There was no difference in accent identification between authentic and imitated accents.

The hypothesis that listeners would not be able to distinguish between authentic and imitated accents was not supported. Listeners were indeed able to determine whether a given voice was authentic or imitated (though the rate itself was not impressive). It seems that, though a strong folk artifact may allow a listener to associate a voice with a certain accent, the folk artifact association does not overpower individual speaker cues that may indicate the authenticity of a speaker.

The final hypothesis was that listeners would react differently to free imitations and modeled imitations. Results for this were mixed. Listeners did identify the accent of free and modeled imitations differently, but there was no difference in identifying the authenticity of free and modeled imitations. Listeners were better at identifying the accent of free imitations than modeled imitations; this leads us to conclude that imitations based strictly on folk artifacts (as opposed to immediate targets) are easier for listeners to recognize. The fact that there was no difference when identifying the authenticity shows that both types of imitation were correctly recognized as imitations.

Listeners' identification of accents varied significantly by language. The accents from most to least often identified were Chinese, German, Spanish, Indian, French. Authenticity identification did not vary by language. The misidentifications did not show any complementariness between languages.

There was a weak correlation between correct identification of accent and authenticity. If listeners got one identification correct, they were very slightly more likely to get the other identification correct as well.

There was a correlation between the ratings by linguists and the identification of accents by listeners: the higher the linguist rating, the more likely the accent was to be identified. There was no such correlation for authenticity identification.

Listener confidence predicted listener accuracy, for both accent and authenticity identifications.

There were two types of listener familiarity, level and method, which were not correlated with each other. Method of familiarity was not significantly related to identification. Level of familiarity was significantly correlated with authenticity identification, but not with accent identification. This shows that a listener's greater familiarity with an accent may influence their ability to discern whether a given voice is authentic or imitated, but it had no impact otherwise.

The following chapter discusses how these listener results reflect on folk artifacts and linguistic awareness.

Chapter 6 Towards a Theory of Folk Artifact Imitation

This dissertation analyzed imitations and folk artifacts from a folk linguistic awareness perspective. This chapter begins by reviewing the hypotheses and how my research both validated and refuted them. I then propose a theory of folk artifact imitation that accounts for these results. The implications for folk linguistic awareness, folk linguistic artifacts, and other linguistic fields are discussed. The primary conclusions are that folk linguistic awareness provides a more detailed understanding of salience and its relationship to imitation, and that folk linguistic artifacts are the basis of imitations and assist in listener categorization. The chapter ends by presenting limitations to this research and suggestions for future research.

6.1 Hypotheses

The hypotheses were developed by addressing perception, and specifically aspects of awareness, from two angles: that of the speaker and that of the listener. Speakers' awareness was revealed in their productions of non-native imitations, which allowed me to investigate whether free or modeled imitations were more similar to the target accent. Listeners' awareness was revealed in their evaluations of these imitations. This allowed me to investigate whether listeners could accurately identify accents, whether they could distinguish between authentic and

imitated accents, and whether they reacted differently to free and modeled imitations. Specifically, the hypotheses were as follows.

1. Speakers' modeled imitations will be more similar to the authentic accents than their free imitations.
2. Listeners will be able to identify the NNAE accents, both authentic and imitated.
3. Listeners will not be able to distinguish between authentic and imitated accents.
4. Listeners will react differently to free imitations and modeled imitations.

The first hypothesis was that *speakers' modeled imitations would be more similar to the authentic accents than their free imitations.* For the features investigated here (pitch, vowel duration, and vowel formants), both free and modeled imitations were closer to the authentic accents than to their own regular guise; however, neither imitation type was closer to the authentic than the other, and therefore this hypothesis was rejected. The rationale behind this first hypothesis was that free imitations were based on folk artifacts, and modeled imitations built on these with additional features from the authentic target accent, so the modeled imitation would be more similar to the authentic accents. This was based on the findings of Torstensson et al. 2004. However, the present data did not show a consistent significant difference between the free and modeled imitations.

The lack of difference between the two types of imitations indicates that both were predominantly based on folk artifacts. This is presumed because the modeled imitations did not show significant changes after exposure to the authentic target

accents; they were very similar to the previously performed free imitations. There are two possible reasons why the presence of authentic targets did not lead to significant changes in the pitch and vowels of the modeled imitations. The first possibility is that the differences between the authentic features and artifact features were large, but the artifact itself was overpowering, and could not be changed significantly by brief exposure to the authentic target. The second possibility is that the differences between the authentic target and the folk artifact were small, so even if the target did influence the artifact, it was not enough to register as a significant change. Either way, the authentic targets were incorporated into the artifacts without causing large vowel or pitch modifications to the imitations. This likely reflects the strength of these particular folk artifacts.

The second hypothesis was that *listeners would be able to identify the accents.* This hypothesis was correct, as listeners indeed recognized the (target) language of both authentic and imitated accents. These specific accents have prominent folk artifacts (as seen in Lindemann 2005 and the pilot study presented in 2.6), which likely assisted with identification. Neuhauser & Simpson 2007 had also found that listeners identified the accents correctly, with imitations being identified better than authentic accents; however in the present study there was no statistical difference in the rates of accent identification between imitations and authentic accents.

The third hypothesis was that *listeners would not be able to distinguish between authentic and imitated accents.* This was incorrect, as listeners were able to discern which voices were authentic and which were imitations, although the rates themselves were not impressive. This result is contradictory to Neuhauser and

Simpson 2007, who found that listeners were unable to distinguish between imitated and authentic accents (though their results were not tested for significance). The original rationale behind this third hypothesis was that the strength of a voice's out-group association with a folk artifact would overpower individual speaker details. However, the present listeners' ability to differentiate between authentic and imitated voices leads us to conclude that the strength of the folk artifact association does not suppress the individual speaker cues that indicate authenticity.

The fourth hypothesis was that *listeners would react differently to free imitations and modeled imitations.* This hypothesis was partially correct. While there was no difference in how the authenticity of these imitations was judged (both types were correctly identified as imitations), there was a significant difference in how well the accents of free and modeled imitations were recognized, such that listeners were better at identifying the accent of free imitations than modeled imitations.[19] In light of the results from the first hypothesis, this outcome is unexpected. Though there were no significant differences between the speakers' free and modeled vowels or pitch, there must have been other differences between the two types of imitations that led listeners to recognize free imitations better than modeled. The imitations based solely on folk artifacts were easier for listeners to recognize than those imitations ostensibly based on immediate authentic targets. This is not what we would have expected if modeled imitations really were just adding on to the free imitations with features from the target, as Torstensson et al. 2004 asserted their

[19] Note that the accents of both types of imitations were correctly identified at a significant rate; this focuses on the difference between the rates, which was significant.

speakers did. It is unclear why the modeled imitation accent identifications are worse then the free. The present analysis of pitch and vowels cannot resolve this, since there were no significant differences between free and modeled imitations. Though the limited phonetic results cannot provide complete details, the listener results nonetheless indicate that when it comes to identifying a speaker's accent, imitations based solely on folk artifacts are easier for listeners to recognize.

Based on the results of this research, we may formulate a theory of folk artifact imitations. I propose that folk linguistic artifacts are the single largest influence on imitations. Free (spontaneous) imitations best reflect these artifacts. Modeled imitations differ slightly from free imitations in their linguistic modifications, due to the presence of immediate authentic target accents, but they are still largely based on folk artifacts. Listeners can identify the accents of the imitations in this research because they are associated with strong folk artifacts. Listeners are better at identifying the accent of free imitations than modeled imitations, because they are more similar to the shared folk artifacts which listeners also hold. This is in spite of the fact that listeners can recognize that these imitations are not authentic.

6.2 Implications

This research began by asking two broad questions: what can imitation reveal about perception, and how are folk linguistic artifacts involved in perception? In response to the first question, we have seen that imitation can reveal aspects of

folk linguistic awareness, in particular which features are controllable. As I will discuss below, this is related to and yet different from which features are salient. In response to the second question, this research suggests that folk artifacts are the foundation of imitations and that they assist in out-group categorization.

We can break these answers down into their implications for folk linguistic awareness and implications for folk linguistic artifacts. I will then discuss the applications this work has for other linguistic fields.

6.2.1 Implications for Folk Linguistic Awareness

As discussed in 1.1.2, folk linguistic awareness is made up of four modes: control, availability, detail, and accuracy. By analyzing imitations we can explore the varying interaction of these modes. The imitation itself is most reflective of the mode of control; it reveals which features speakers can modify. The modifications may be more or less accurate and more or less detailed. Imitation can also reflect availability to an extent: a feature must be available on some level in order for it to be modified.

The fact that imitations can reflect availability seems to be what has led some linguists to suggest that imitations reflect *salience* (e.g. Trudgill 1986, Schilling-Estes 1995). Definitions of salience usually refer to awareness; by using the modes of folk linguistic awareness, we can more precisely define salience as *the available details.* Notice that this definition does not refer to the mode of control. Imitation can indicate salience, as it is likely that features modified in imitations highlight the

details that are more available, but that does not mean that other features, which are less easily controlled and therefore not modified in imitations, are not also salient. Additionally, the imitated features are not necessarily accurate representations of salient features: if a feature is available and yet difficult to control, it may be modified but in a manner which is different from how it is perceived. Gross changes may also mask specifics of the details. My theory of folk artifact imitation predicts that features which are modified in imitations are a good place to start when investigating salience, but they are ultimately more reflective of control than availability, accuracy, or detail.

The features which were modified in imitations of NNAE accents in this study are the following: mean pitch, vowel duration, F1 (particularly of /a, E, I, ou, ow/), and F2 (particularly of /a, ae, aw, ou, ow, uw, I/). The fact that these features were consistently modified by the speakers indicates that they are controllable. That these modifications were controlled would not be enough in itself to assert that the imitated features were accurate; however, their similarity in this case to authentic accents is evidence that they are fairly accurate. That these features were modified does indicate that they are available, and therefore likely to be salient. While the lack of modifications of other features does not necessarily mean those features are not also salient, an investigation of salient features of NNAE accents would do well to begin with those which were modified in this study.

6.2.2 Implications for Folk Linguistic Artifacts

Folk linguistic artifacts are conventionalized conceptual representations of language varieties. Artifacts are essentially cognitive templates that may vary in strength; for example, a German artifact may be strong while a Romanian artifact may be weak. Many artifacts have a geographic basis. Artifacts may be specific or broad, and they may overlap; for example, a Chinese artifact can specifically refer to a Mandarin Chinese accent and also broadly refer to an Asian accent.

In this study, free imitations were elicited to display folk artifacts. Modeled imitations were intended to manifest the influence of immediate targets. Interestingly, there were no significant phonetic differences between the two types of imitations in measures of pitch, vowel duration, or vowel formants. Because of this, both types of imitations are quite similar, leading us to conclude that both are predominantly based on folk artifacts. Because the free imitations were elicited first, and the modeled imitations did not show significant differences from them, the underlying basis of both types of imitations (at least for the specific features investigated) is assumed to be the folk artifacts that informed the initial free imitation. This data shows, for the first time, what a NNAE folk artifact consists of phonetically: speakers who are imitating NNAE accents tend to raise their mean pitch, exaggerate the length difference between monophthongs and diphthongs, slightly raise some vowels, and back the mid and back vowels. These modifications produce imitations which are more similar in these measurements to authentic accents than to the imitators' regular voices. Both free imitations and modeled imitations share these features.

Despite the lack of vowel or pitch differences between the free and modeled imitations, listeners did identify the accents of the two types of imitations differently, indicating that there were other linguistic differences between the free and modeled imitations and justifying maintaining a distinction between the two. Listeners were better at identifying the accents of free imitations, suggesting that listeners prefer imitations that match their own artifacts, rather than imitations that are more explicitly modeled after authentic accents (though without the relevant phonetic measurements we cannot say to what extent these were modeled). Listeners can more easily categorize out-group speech when it is based entirely on folk artifacts (as in the free imitations) than when it is also based on authentic out-group speech (as in the modeled imitations).

The effect of familiarity on folk artifacts appears ambiguous. In particular, the lack of relationship found here between listener familiarity and the identification of accents is puzzling. I had initially assumed that familiarity would influence identification, based in part on Preston's assertion that both face-to-face interaction and media exposure are ways that artifacts are transmitted (1996:59). Familiarity is often identified as a significant factor in sociolinguistic studies (e.g. Derwing & Munro 1997; Clopper & Pisoni 2004b), and occasionally familiarity through media exposure is specifically cited (e.g. Trudgill 1983; Kerswill & Williams 2002). However, these studies did not specifically test for an effect of type of familiarity, and in fact several of them did not actually measure an amount of familiarity. In the present research, the level (that is, the amount) of familiarity with an accent was not correlated with accent identification, though it did have a positive impact on

authenticity identification: the more familiar a listener was with an accent, the more likely they were to correctly identify the authenticity of a speaker. The familiarity of a listener did not affect how they identified the accent of a speaker, however. There was also a lack of significant effect of method (that is, type) of familiarity: notably, there were no significant differences in identification between people who were familiar with accents through personal (face-to-face) interaction, through media exposure, and through language study. I propose that familiarity was not significantly correlated with accent identification because the folk artifacts used are so ingrained. Familiarity plays less of a role in distinguishing between the strongest folk artifacts, because they are all very well known. My theory of folk artifact imitation predicts that familiarity (particularly through the media and through personal experience) would play a larger role in the accent identification of weaker artifacts, because familiarity would vary more between listeners. Listeners would not be able to identify an accent which they did not have an artifact for. Both familiarity and folk artifact strength should be carefully evaluated to determine their influence on listeners' identification.

6.2.3 Implications for Other Fields

The theory of folk artifact imitation I have put forward in the preceding sections is also applicable to other linguistic fields.

Sociolinguists who have concentrated on production in language variation must also study perception to ensure that the features they have connected with identity are perceived that way by listeners. My research suggests that there is an

interaction between those linguistic features that are most often changed (or most often held onto and exaggerated) and those that are most salient; a theory of folk artifact imitation and the folk linguistic awareness framework can help sociolinguists to interpret this. The present results should encourage sociolinguists to embrace studies involving imitation (and more broadly, performance), as an opportunity to take a different approach to understanding how varieties are perceived. It is possible for imitations to reveal more than gross stereotypes; as this study showed, imitations can reflect specific phonetic features of the target variety, such as pitch and vowel formants, with appreciable accuracy.

The field of speech perception would also benefit from using imitation as a method to identify potential perceptual cues. A more in-depth understanding of salience would also be beneficial in this realm. By studying folk linguistic artifacts, linguists could better understand individuals' cognitive templates and how artifacts might assist the process of categorization.

This research is also applicable to the field of forensic linguistics. Forensic linguists are interested in knowing which features are more or less controllable, as this may aid them in investigating cases of voice disguise where speakers manipulate their speech. By studying which specific features are modified in experimental imitations, forensic linguists may discover which specific features are being modified in actual instances of disguised imitations.

Finally, the perception of accented speech is not just an interesting research question; it is a topic with important real-world consequences. Purnell et al. 1999 discovered that housing discrimination occurs based on recognition of African

American and Hispanic accents. Baugh 2000 labels this phenomenon *linguistic profiling* and states it is "based upon auditory cues that may be used to identify an individual or individuals as belonging to a linguistic subgroup within a given speech community" (363). The present research shows that listeners are able to identify NNAE accents, and the risk of NNAE speakers being profiled and subsequently treated unfairly is alarming.

6.3 Future Research

This study was able to avoid the more common difficulties in general imitation research of conflating the modes of linguistic awareness, not clarifying the target, and misinterpreting the modifications (c.f. 1.3.2), though it did have its own limitations. Some of these limitations lead to suggestions for future research.

The first common limitation of much imitation research is that the modes of linguistic awareness are not distinguished. Researchers mistakenly assume that high control is synonymous with high availability, and this is how the *imitation = salience* assumption occurs. By recognizing that the imitations reflect primarily the mode of control, and acknowledging that this is related to yet distinct from the other modes of availability, accuracy, and detail, I was able to avoid this difficulty.

Another common difficulty with imitation research is that the target of the imitation is unclear. This study avoided this limitation by distinguishing between the targets of free imitations, which were folk artifacts held by speakers, and the targets of modeled imitations, which were authentic voices heard by speakers.

The final general difficulty of imitation research is in interpreting the linguistic modifications in imitations. It cannot be automatically assumed that imitations do or do not reflect the target accurately. By analyzing speakers' regular voices in addition to their imitations, and comparing these varieties to the authentic target voices, I could determine to what extent the modifications accurately reflected the authentic target.

One limitation of this particular study was the number of different speakers: a total of four per language, with two authentic speakers and two imitation speakers each. Logistic constraints necessitated this amount, in order to keep the survey to a reasonable time for listeners to complete. However, most listeners did not need to listen to a speaker for an entire minute in order to make their judgment, and they finished the survey rapidly. Future research could present much less speech from each speaker (perhaps 15 seconds instead of 60 seconds) and therefore be able to include a greater number of total speakers. A larger number of speakers would ensure that individual speaker idiosyncrasies would not unduly influence the final results.

Listeners can and do use many different cues in perceiving speech. Another limitation of this study, therefore, is that only three linguistic cues were analyzed: pitch, vowel duration, and vowel formants. The contribution of other features is very important, and future research should analyze both additional segmental and suprasegmental cues. Once these features have been investigated, further studies may focus on those particular features which are significant in order to determine more precisely how they are perceived and modified. For example, future studies

might focus on individual artifacts to determine which features listeners are most aware of for that specific artifact.

To determine more definitively whether listeners perceive differences between free and modeled imitations, the two types of imitations may be presented directly against each other (either as a whole or broken down into individual cues), with listeners being asked explicit questions to compare the speech, such as which sounds more authentic, which sounds more like a e.g. French accent, or simply if they are different. This would allow us to verify if listeners do perceive the differences between imitations with different targets, and potentially which phonetic features assist in this.

This dissertation supports the presence of folk artifacts. Further studies could elaborate on how those artifacts are categorized, perhaps in an auditory free classification task (Clopper & Pisoni 2007, Clopper 2008) designed to test the groupings found in Lindemann 2005. This type of experiment could also shed light on the distinction between broad and specific artifacts.

Additional research may also be done to investigate how folk linguistic artifacts are created and transmitted. While it is likely that American artifacts of NNAE speech are formed from a combination of personal experience and media exposure, the present results do not indicate a straightforward relationship between familiarity and accent identification. A more nuanced investigation of the influence of familiarity on folk artifact development and identification would be valuable.

6.4 Conclusion

This study has examined the perception of imitations in the light of folk linguistic concepts. Folk linguistic awareness, with its four modes of control, availability, detail, and accuracy, has been shown to be a valuable approach to interpreting imitations. This type of awareness is also important for better understanding the concept of salience. A theory of folk artifact imitation identifies folk linguistic artifacts as the basis of imitations and an important tool in categorizing speech. Overall, imitation revealed aspects of perception, confirming its importance and validity as a linguistic research tool.

Bibliography

Abrams, Dominic & Michael A. Hogg. 1987. Language attitudes, frames of reference, and social identity: a Scottish dimension. *Journal of Language and Social Psychology* 6:201-213.

Baker, Wendy, David Eddington & Lyndsey Nay. 2009. Dialect identification: the effects of region of origin and amount of experience. *American Speech* 84(1):48-71.

Baugh, John. 2000. Racial identification by speech. *American Speech* 75(4):362-364.

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

Boersma, Paul & Weenink, David. 2008. Praat: doing phonetics by computer (Version 5.0.45). [Computer program: http://www.praat.org/]

Brunner, Elizabeth. 2008. The identification of foreign-accented English: a sociolinguistic topic with forensic significance. NWAV 37, Houston, November 6-9, 2008.

Bush, Clara N. 1967. Some acoustic parameters of speech and their relationships to the perception of dialect differences. *TESOL Quarterly* 1(3):20-30.

Bucholtz, Mary. 2003. Sociolinguistic nostalgia and the authentication of identity. *Journal of Sociolinguistics* 7(3):398-416.

Chance, June E. & Alvin G. Goldstein. 1996. The other-race effect and eyewitness identification. In Siegfried Ludwig Sporer, Roy S. Malpass & Guenter Koehnken (Eds) *Psychological Issues in Eyewitness Identification.* Mahwah, NJ: Lawrence Erlbaum Associates. 153-176.

Chun, Elaine. 2006. Taking the mike: performances of everyday identities and ideologies at a U.S. high school. *Texas Linguistic Forum* 49:39-49.

Clopper, Cynthia G. 2008. Auditory free classification: methods and analysis. *Behavior Research Methods* 40(2):575-581.

Clopper, Cynthia G. & David B. Pisoni. 2004a. Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics* 32:111-140.

Clopper, Cynthia G. & David B. Pisoni. 2004b. Homebodies and army brats: some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change* 16:31-48.

Clopper, Cynthia G. & David B. Pisoni. 2006. Effects of region of origin and geographic mobility on perceptual dialect categorization. *Language Variation and Change* 18:193-221.

Clopper, Cynthia G. & David B. Pisoni. 2007. Free classification of regional dialects of American English. *Journal of Phonetics* 35:421-438.

Clopper, Cynthia G. & David B. Pisoni. 2004. Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics* 32:111-140.

Coupland, Nikolas. 2003. Sociolinguistic authenticities. *Journal of Sociolinguistics* 7(3):417-431.

Coupland, Nikolas. 2007. *Style: Language Variation and Identity.* Cambridge: Cambridge University Press.

Cunningham-Andersson, Una. 1996. Learning to interpret sociodialectal cues. *TMH-QPSR* 37(2):155-158.

Derwing, Tracey M. & Munro, Murray J. 1997. Accent, intelligibility, and comprehensibility: evidence from four L1s. *Studies in Second Language Acquisition* 20: 1--16.

Doty, Nathan Daniel. 1998. The influence of nationality on the accuracy of face and voice recognition. *The American Journal of Psychology* 111:191--214.

Drager, Katie. 2006. From *bad* to *bed*: the relationship between perceived age and vowel perception in New Zealand English. *Te Reo* 48:55-68.

Duszak, Anna (Ed). 2002. *Us and Others: Social Identities Across Languages, Discourses, and Cultures.* Philadelphia: John Benjamins.

Eckert, Penelope. 2003. Elephants in the room. *Journal of Sociolinguistics* 7(3):392-397.

Evans, Betsy E. 2002. An acoustic and perceptual analysis of imitation. In Daniel Long & Dennis R. Preston (Eds) *The Handbook of Perceptual Dialectology, Volume 2.* Philadelphia: John Benjamins. 95-112.

Flege, James Emil. 1984. The detection of French accent by American listeners. *Journal of the Acoustical Society of America* 76(3):692-707.

Flege, James E. & Robert M. Hammond. 1982. Mimicry of Non-distinctive Phonetic Differences Between Language Varieties. *Studies in Second Language Acquisition* 5(1): 1-17.

Giles, Howard (Ed). 1977. *Language, Ethnicity, and Intergroup Relations.* London: Academic Press.

Hammersley, Richard & J. Don Read. 1996. Voice identification by humans and computers. In Siegfried L. Sporer, Roy S. Malpass & Guenter Koehnken (Eds) *Psychological Issues in Eyewitness Identification.* Mahwah, NJ: Lawrence Erlbaum. 117-152.

Hay, Jennifer, Paul Warren & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34:458-484.

Hickey, Raymond. 2000. Salience, stigma and standard. In Laura Wright (Ed) *The Development of Standard English 1300-1800.* Cambridge: Cambridge University Press. 57-72.

Ikeno, Ayako. 2005. *Perceptual Cues in English Accent Variation: The Role of Prosody and Listener Accent Background.* PhD Dissertation, University of Colorado.

Johnson, Keith, Elizabeth A. Strand & Mariapaola D'Imperio. 1999. Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics* 27:359-384.

Kang, Okim & Donald L. Rubin. 2009. Reverse linguistic stereotyping: measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology* 28(4):441-456.

Kerstholt, Jose H., Noortje J.M. Jansen, Adri G. van Amelsvoort, & A.P.A. Broeders. 2006. Earwitnesses: effects of accent, retention and telephone. *Applied Cognitive Psychology* 20:187-197.

Kerswill, Paul & Ann Williams. 2002. Dialect recognition and speech community focusing in new and old towns in England: The effects of dialect levelling, demography and social networks. In Daniel Long & Dennis R. Preston (Eds) *The Handbook of Perceptual Dialectology, Volume 2.* Philadelphia: John Benjamins. 173-204.

Koops, Christian, Elizabeth Gentry & Andrew Pantos. 2008. The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics* 14(2):91-101.

Künzel, Hermann J. 2000. Effects of voice disguise on speaking fundamental frequency. *International Journal of Speech, Language, and the Law* 7(2): 149-179.

Labov, William. 1966. *The Social Stratification of English in New York City.* Washington, D.C.: Center for Applied Linguistics.

Labov, William. 1972. Sociolinguistic Patterns. Philadelphia: University of Pennsylvania.

Labov, William. 2001. *Principles of Linguistic Change, Vol 2: Social Factors.* Oxford: Blackwell.

Labov, William, Paul Cohen, Clarence Robins, & John Lewis. 1968. A study of the non-standard English of Negro and Puerto Rican speakers in New York City. Final report, Cooperative Research Project 3288. 2 vols. Philadelphia: U.S. Regional Survey.

Lambert, Wallace E., Moshe Anisfeld & Grace Yeni-Komshian. 1965. Evaluational reactions of Jewish and Arab adolescents to dialect and language variations. *Journal of Personality and Social Psychology* 2:84-90.

Lambert, Wallace E. 1967. A social psychology of bilingualism. *Journal of Social Issues* 23(91-109).

Lindemann, Stephanie. 2003. Koreans, Chinese or Indians? Attitudes and ideologies about non-native English speakers in the United States. *Journal of Sociolinguistics* 7(3): 348-364.

Lindemann, Stephanie. 2005. Who speaks 'broken English'? US undergraduates' perceptions of non-native English. *International Journal of Applied Linguistics* 15(2): 187--212.

Lindsey, Geoff & Allen Hirson. 1999. Variable Robustness of Nonstandard /r/ in English: Evidence from Accent Disguise. *Forensic Linguistics* 6(2): 278-288.

Daniel Long & Dennis R. Preston (Eds). 2002. *The Handbook of Perceptual Dialectology, Volume 2.* Philadelphia: John Benjamins.

Markham, Duncan. 1997. Phonetic imitation, accent, and the learner. *Travaux de l'Institut de Linguistique de Lund* 33:i-269.

Markham, Duncan. 1999. Listeners and disguised voices: The imitation and perception of dialectal accent. *Forensic Linguistics* 6(2):289-299.

Masthoff, Herbert. 1996. A Report on a Voice Disguise Experiment. *Forensic Linguistics* 3(1): 160-167.

Munro, Murray J. 1995. Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition* 17:17-34.

National Geographic-Roper Public Affairs. 2006. Geographic Literacy Study. New York, NY: GfK NOP. [Available at http://www.nationalgeographic.com/roper2006/]

Neuhauser, Sara & Adrain P. Simpson. 2007. Imitated or authentic? Listeners' judgments of foreign accents. Conference proceedings of ICPhS XVI, Saarbrucken August 6-10, 2007. 1805-1808.

Niedzielski, Nancy A. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18(1):62-85.

Niedzielski, Nancy A. & Dennis R. Preston. 2000. *Folk Linguistics.* New York: Mouton de Gruyter.

Podbresky, Rosita, Robert H. Deluty & Stanley Feldstein. 1990. Evaluations of Spanish- and Oriental-accented English speakers. *Social Behavior and Personality* 18(1):53-63.

Preston, Dennis R. 1989. Perceptual Dialectology: Nonlinguists' Views of Areal Linguistics. Providence, RI: Foris Publications.

Preston, Dennis R. 1992. Talking Black and talking White: A study in variety imitation. In Joan H. Hall, Nick Doane & Dick Ringler (Eds) *Old English and New: Studies in Language and Linguistics in Honor of Frederic G. Cassidy.* New York: Garland Publishing. 327-355.

Preston, Dennis R. 1996. Whaddayaknow? The modes of folk linguistic awareness. *Language Awareness* 5(1):40-74.

Preston, Dennis R. 1999. A language attitude approach to the perception of regional variety. In Dennis R. Preston (Ed) *Handbook of Perceptual Dialectology, Volume 1.* Philadelphia: John Benjamins. 359-373.

Purnell, Thomas, William Idsardi & John Baugh. 1999. Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology* 18(1):10-30.

Schiller, Niels O. & Olaf Köster. 1998. The ability of expert witnesses to identify voices: a comparison between trained and untrained listeners. *Forensic Linguistics* 5(1):1-9.

Schilling-Estes, Natalie. 1995. Production, perception and patterning: "Performance" speech in an endangered dialect variety. *Penn Working Papers in Linguistics* 2(2):117-131.

Schilling-Estes, Natalie. 1998. Investigating "self-conscious" speech: The performance register in Ocracoke English. *Language in Society* 27:53-83.

Schlichting, Frank & Kirk P.H. Sullivan. 1997. The Imitated Voice – a Problem for Voice Line-ups? *Forensic Linguistics* 4(1): 18-165.

Strand, Elizabeth A. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18(1):86-99.

Tajfel, Henri (Ed). 1978. *Differentiation Between Social Groups: Studies in the Social Psychology of Intergroup Relations.* London: Academic Press.

Tate, Donna A. 1977. Preliminary data on dialect in speech disguise. *Current issues in the phonetic sciences: Proceedings of the IPS-77 Congress* 9. Amsterdam: Benjamins. 847-850.

Thomas, Erik R. 2002. Sociophonetic applications of speech perception experiments. *American Speech* 77(2):115-147.

Thomas, Erik R. & Tyler Kendall. 2007. NORM: The vowel normalization and plotting suite. [Online Resource: http://ncslaap.lib.ncsu.edu/tools/norm/]

Thomas, Erik R. & Jeffrey Reaser. 2004. Delimiting perceptual cues used for the ethnic labeling of African American and European American voices. *Journal of Sociolinguistics* 8(1):54-87.

Thompson, Charles P. 1987. A language effect in voice identification. *Applied Cognitive Psychology* 1:121-131.

Torbert, Benjamin. 2004. *Southern Vowels and the Social Construction of Salience.* PhD dissertation, Duke University.

Torstensson, Niklas, Erik J. Eriksson, & Kirk P.H. Sullivan. 2004. Mimicked Accents – Do Speakers Have Similar Cognitive Prototypes? Proceedings of the 10th Australian International Conference on Speech Science & Technology, Macquarie University, Sydney, December 8-10, 2004. 271-276.

Trudgill, Peter. 1983. *On Dialect: Social and Geographical Perspectives.* New York: New York University Press.

Trudgill, Peter. 1986. *Dialects in Contact.* Oxford: Basil Blackwell.

Wertheim, Suzanne. 2003. Rethinking the observer's paradox and data "purity." In J. Larson and M. Paster (Eds) *Proceedings of the 28th Meeting of the Berkeley Linguistics Society*. Berkeley: BLS. 511-521.

Williams, Angie, Peter Garrett & Nikolas Coupland. 1999. Dialect recognition. In Dennis R. Preston (Ed) *Handbook of Perceptual Dialectology, Volume 1*. Philadelphia: John Benjamins. 345-358.

Wretling, Pär & Anders Eriksson. 1998. Is articulatory timing speaker specific? Evidence from imitated voices. *Proceedings of FONETIK 98.* 48-52.

Zetterholm, Elisabeth. 1997. Impersonation: a phonetic case study of the imitation of a voice. *Lund University Department of Linguistics Working Papers* 46:269-287.

Zetterholm, Elisabeth. 2003. Voice imitation: a phonetic study of perceptual illusions and acoustic success. *Travaux de l'Institut de Linguistique de Lund* 44:i-149.

Zuengler, Jane. 1988. Identity markers and L2 pronunciation. *Studies in Second Language Acquisition* 10:33-49.

Appendix A

Reading Passage

Listen carefully Mr. Doe,

I know all about what you've been doing. You've cheated hundreds of people by making them pay for things you never gave them. You thought just because they're old that you could take advantage of them and they wouldn't figure it out. Bad choice. I'm here to say that not everyone believes your lies. Now it's time for you to pay your dues.

You are going to give the money back to all those people. Then you're going to pay me for showing you the error of your ways.

Put ninety thousand dollars in cash in that white gym bag you have. That's exactly the amount you stole, plus five thousand for me. On Tuesday morning at 10 a.m., leave the bag with the book vendor on the south-east corner of Hitch and Zane streets.

This is not a joke. I don't think your daughter Dawn would be proud of her dad if she knew what he did.

Don't bother telling the police – that would just make things worse for you. I mean it. No cops or the deal is off.

You can't stop me. Ninety thousand in cash, Tuesday at ten – or else.

Appendix B

Folk tale passages read between imitations

I. Goldilocks and the Three Bears (read between free imitations)

Passage 1

Once upon a time, there was a little girl named Goldilocks. She went for a walk in the forest. Soon she came upon a house. She knocked and, when no one answered, she walked right in.

At the table in the kitchen, there were three bowls of porridge. Goldilocks was hungry. She tasted the porridge from the first bowl.

"This porridge is too hot!" she exclaimed.

So, she tasted the porridge from the second bowl.

"This porridge is too cold," she said

So, she tasted the last bowl of porridge.

"Ahhh, this porridge is just right," she said happily and she ate it all up.

Passage 2

After Goldilocks had eaten the three bears' breakfasts she decided she was feeling a little tired. So, she walked into the living room where she saw three chairs. Goldilocks sat in the first chair to rest her feet.

"This chair is too big!" she exclaimed.

So she sat in the second chair.

"This chair is too big, too!" she whined.

So she tried the last and smallest chair.

"Ahhh, this chair is just right," she sighed. But just as she settled down into the chair to rest, it broke into pieces!

Passage 3

Goldilocks was very tired by this time, so she went upstairs to the bedroom. She lay down in the first bed, but it was too hard. Then she lay in the second bed, but it was too soft. Then she lay down in the third bed and it was just right. Goldilocks fell asleep.

As she was sleeping, the three bears came home.

"Someone's been eating my porridge," growled Papa Bear.

"Someone's been eating my porridge," said Mama Bear.

"Someone's been eating my porridge and they ate it all up!" cried Baby Bear.

They went from the kitchen to the living room.

"Someone's been sitting in my chair," growled Papa Bear.

"Someone's been sitting in my chair," said Mama Bear.

"Someone's been sitting in my chair and they've broken it all to pieces," cried Baby Bear.

Passage 4

They decided to look around some more and when they got upstairs to the bedroom, Papa Bear growled, "Someone's been sleeping in my bed."

"Someone's been sleeping in my bed, too" said Mama Bear

"Someone's been sleeping in my bed and she's still there!" exclaimed Baby Bear.

Just then, Goldilocks woke up and saw the three bears. She screamed and jumped up. Goldilocks ran down the stairs, opened the door, and ran away into the forest. And she never returned to the home of the three bears.

II. The Little Boy Who Cried Wolf (read between modeled imitations)

Passage 1

There was once a poor shepherd boy who used to watch his flocks in the fields next to a dark forest near the foot of a mountain. One hot afternoon, he thought up a good plan to get some company for himself and also have a little fun. Raising his fist in the air, he ran down to the village shouting "Wolf, Wolf!"

Passage 2

As soon as the townspeople heard the shepherd boy, they all rushed from their homes, full of concern for his safety, and two of his cousins even stayed with him for a short while. This gave the boy so much pleasure that a few days later he tried exactly the same trick again, and once more he was successful.

Passage 3

However, not long after, a wolf that had just escaped from the zoo was looking for a change from its usual diet of chicken and duck. So, overcoming its fear of being shot, it actually did come out from the forest and began to threaten the sheep.

Passage 4

Racing down to the village, the shepherd boy of course cried out even louder than before. Unfortunately, as all the villagers were convinced that he was trying to fool them a third time, they told him, "Go away and don't bother us again." And so the wolf had a feast.

Appendix C

Vowel space by language

A normalized vowel space of all speakers, separated by *language*. Each language represents the two NNAE speakers and four imitations (a free and a modeled, from two speakers) of the language; "Regular (AE)" represents all ten regular AE voices.